TITLE

The Trade-Offs of Multicast Routing Protocols

DEFENCE DÉFENSE

The Trade-Offs of Multicast Routing Protocols

Claude Bilodeau
Communications Research Centre Canada

Communications Research Centre
Centre de recherches sur les communications

The work described in this document was sponsored by the Department of National Defence under Work Unit 5CB14

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-119
CRC Report No. 99-004
January 1999

National Defence Défense nationale

Canada

DEFENCE DÉFENSE

The Trade-Offs of Multicast Routing Protocols

Claude Bilodeau
Broadband Network Technologies Branch
Communications Research Centre Canada

Communications Research Centre
Centre de recherches sur les communications

The work described in this document was sponsored by the Department of National Defence under Work Unit 5CB14

DEFENCE RESEARCH ESTABLISHMENT OTTAWA

TECHNICAL REPORT
DREO TR 1999-119
CRC Report No. 99-004
January 1999

Project
5CB14

Abstract

During the past few years, several multicast routing protocols have emerged, which are competing to provide efficient mechanisms to deliver Internet Protocol (IP) traffic to groups of users scattered throughout the Internet. The multiplicity of experimental protocols and the absence of any well-established standardised protocol for multicast routing indicates that multicast routing has many solutions and that no one implementation can provide the most satisfactory characteristics in every situation.

This paper shows that much work is still needed to advance the state of the multicast routing technology. The main deficiencies of multicast routing protocols and their challenging design issues are illustrated by focusing on a few of the most popular multicast protocols being designed or experimented with today by the Internet Engineering Task Force (IETF).

Most of the multicast routing technology trade-offs analysed in the report apply to the global Internet in general while some are more specific to the tactical communication networks.

Résumé

Ces dernières années, plusieurs protocoles de routage multipoint sont apparus, chacun d'eux procurant des mécanismes efficaces pour acheminer l'information IP (Internet Protocol) vers les groupes participants répartis à travers le réseau Internet. La multiplicité des protocoles expérimentaux et l'absence de protocoles normalisés et bien établis indiquent, somme toute, que le routage multipoint a plusieurs solutions et qu'une mise en oeuvre donnée ne peut exceller et suffire à elle seule à combler les besoins des nombreuses configurations existantes ou futures.

Ce rapport note que beaucoup d'efforts seront nécessaires pour faire avancer la technologie du routage multipoint. Les lacunes principales, et les défis sous-jacents à la conception des protocoles de routage multipoint, sont présentés en examinant quelques-uns des protocoles les plus populaires, certains étant toujours en cours de développement par l'Internet Engineering Task Force (IETF).

La plupart des compromis techniques mentionnés dans ce rapport sont applicables à tout le réseau Internet, alors que d'autres sont plus spécifiquement applicables aux réseaux de communications tactiques.

Executive Summary

DREO TR 1999-119, CRC-RP-99-004
The Trade-Offs of Multicast Routing Protocols
Claude Bilodeau
CRC, VPNT-RNS, Ottawa

Background

IP multicast is a technique used to provide efficient delivery of IP traffic to groups of users scattered throughout the Internet. It enables many new types of applications and reduces network loads. Examples of these applications include distribution of software updates, propagation of realtime data, efficient network news delivery, distance learning classes, video conferences and distributed interactive simulation. Many of these applications have strict requirements in terms of group membership dynamics, group sender populations, group join latency, etc.

In general, the benefits of IP multicast are undeniable and its use more widespread today than just a few years ago. However, it is also becoming obvious that despite broad industry backing and the support of many vendors of network infrastructure elements (e.g. routers, switches, network interface cards and application software), the definition of the IP multicast architecture lags behind the technology. This is reflected in the many transitional approaches to IP multicast being proposed at the moment, which all attempt to address, with varying degree of success, the most urgent, short-term needs.

The Study

This study recognizes that multicast routing has many solutions and that no one implementation can provide the most satisfactory characteristics in every situation. It shows that further advances must be made in several areas and much experimental work remains to be done before an Internet-wide deployment becomes truly functional. The study observes that much of the uniqueness found in the plethora of new multicast protocols depends on four closely related properties: *hierarchism, scalability, autonomy and Policy/Quality-of-Service (QoS) compliance*. While most protocols support the first property and several the second, very few support the third and fourth properties. Such deficiencies of the existing multicast routing infrastructure restrict the use of multicast applications. It is desirable that the multicast routing infrastructure support all four properties to their fullest extent possible if multicasting in an Internet of ever increasing size and heterogeneity is to be widely available, efficient and optimum.

Until recently, the Internet community has been reluctant to invest in comprehensive Policy/QoS routing, in part due to its complexity. This is about to change. The telecommunications industry has reached a turning point where the IP technology is now ubiquitous and tremendeous efforts are being put into developing a QoS solution to reap the profits of a truly integrated voice and data network. It is likely that the availability of multicasting will only occur on a wide scale once the current Policy/QoS impairments have been removed, i.e. once a solution for guaranteeing the QoS levels needed by the dominant applications (e.g. the emerging telephony applications) over most of the Internet has been found.

The Internet Engineering Task Force (IETF) has yet to converge on Internet standards for both inter-domain and intra-domain multicast routing. A fully IP multicast-enabled Internet requires an inter-domain multicast routing standard protocol that permits some degree of multicast routing autonomy. Current protocols are not designed for multiple autonomous systems and cannot limit the propagation of routing information based on policies and rules that administrators might want to use. The growth of IP multicast is severly limited if all routers must contain all routing information for the entire network. The only way to hide information is with a hierarchical routing topology. For intra-domain multicast routing, two standards —one for dense mode and one for sparse mode— may be necessary to accommodate the full range of multicast applications.

Integration of satellite networks with the Internet is forthcoming. Internet Service Providers relying on cable and satellite communications infrastructures are starting to build new business models and introduce value-added services that include IP multicast. Existing Internet unicast and multicast routing protocols have been designed for optimum performance assuming bidirectional symmetrical communication links. Proposed solutions such as a back channel through tunnelling are sub-optimal and should be used in the interim only.

Military Significance

No existing multicast routing protocol is likely to perform satisfactorily in a military environment. None of them was specifically designed for the wireless and low bandwidth environment that is prevalent in military networks, where link symmetry cannot always be achieved despite being needed for proper operation. Furthermore, most multicast routing protocols require better robustness, adaptability and reliability characteristics to operate in such networks.

Suggestions for Future Research Work

Some of the open issues that require further study include:

i. the ability to send traffic to selected destinations according to some well-defined Policy and with agreed-upon guarantee of provisioned Quality-of-Service (QoS) levels;

ii. the ability to limit the growth of the routing information heap while making the multicast Policy/QoS-based routing work across multiple autonomous systems;

iii. the ability to limit the growth of the routing information heap while making the multicast Policy/QoS-based routing work within a partitioned autonomous system;

iv. the ability to accommodate a wide range of heterogeneous networks, including some low bit rate and unidirectional links.

Sommaire

DREO TR 1999-119, CRC-RP-99-004

Les protocoles de routage multipoint et leurs compromis (en anglais)

Claude Bilodeau
CRC, VPNT-RNS, Ottawa

Introduction

Le routage multipoint est une technique permettant l'acheminement efficace de l'information IP vers les groupes participants répartis à travers le réseau Internet. Il rend possible l'existence de nombreuses applications inédites tout en réduisant les charges du réseau. Parmi ces applications, mentionnons, entre autres, la distribution systématique de mises à jour de logiciels, la dissémination de données en temps réel, l'acheminement efficace des nouvelles des cybergroupes, les classes de téléenseignement, la visioconférence et les applications de simulations interactives distribuées. Plusieurs de ces applications ont des exigences rigoureuses en matière de dynamique d'adhésion au groupe, de composition des groupes émetteurs, des temps de latence à l'adhésion au groupe, etc.

Il est généralement reconnu que le multipoint offre des avantages indéniables. Son usage est aussi plus répandu aujourd'hui qu'il ne l'était il y a quelques années. Cependant, même si le multipoint est approuvé par l'industrie en général et soutenu par les fabricants et vendeurs d'éléments d'infrastructure de réseau (e.g. routeurs, multiplexeurs, cartes d'interface de réseau, logiciel d'application), il semble que la définition d'une architecture multipoint pour le trafic IP demeure en retard sur le développement technologique actuel. Cela se constate par les nombreuses approches transitionnelles couramment proposées, toutes, les unes comme les autres, essayant de solutionner, avec plus ou moins de succès, les problèmes les plus urgents et de répondre aux besoins les plus immédiats.

L'étude

Cette étude souligne que le routage multipoint a plusieurs solutions et qu'une mise en oeuvre donnée ne peut exceller et suffire à elle seule à combler les besoins des nombreuses configurations existantes ou futures. Elle démontre que plusieurs lacunes devront disparaître et que beaucoup d'efforts expérimentaux seront nécessaires avant de pouvoir déployer des applications multipoint vraiment fonctionnelles à la grandeur du réseau Internet. L'étude fait observer que l'unicité que lon retrouve dans la panoplie des nouveaux protocoles de routage multipoint relève de quatre propriétés étroitement liées: division hiérarchique, échelonnabilité, autonomie et conformité aux politiques/qualité-de-services (QoS). La plupart des protocoles adhèrent à la première propriété, plusieurs incluent la seconde, mais très peu considèrent la troisième et la quatrième. De telles déficiences réduisent considérablement la portée des applications multipoint. Pour que le multipoint devienne efficace, optimum et disponible d'un bout à l'autre d'un Internet toujours grandissant et de plus en plus hétérogène, il est souhaitable que l'infrastructure de routage multipoint puisse supporter dans sa totalité les quatre propriétés mentionnées.

En partie à cause de sa complexité, la communauté Internet avait, jusqu'à tout récemment, hésité à investir dans le développement d'un routage subordonné à des règles politiques/QoS. Les choses sont en train de changer. L'industrie des télécommunications a amorcé un virage où la technologie Internet devient omniprésente et où des efforts colossaux sont déployés pour trouver une solution au routage politique/QoS, espérant ainsi tirer de généreux profits par la mise en place de réseaux où données et voix sont vraiment intégrées. Il est à prévoir que le développement du multipoint se fera à grande échelle seulement si les déficiences du routage politique/QoS sont écartées i.e. lorsqu'on aura trouvé une solution qui garantit, à travers l'Internet, les niveaux de QoS nécessaires pour soutenir les applications dominantes (e.g. les toutes nouvelles applications de téléphonie).

L'Internet Engineering Task Force (IETF) se tarde d'adopter des normes pour le routage multipoint intra-domaine et inter-domaines. L'Internet a besoin d'un protocole de routage multipoint qui normalise mais aussi tolère un certain degré d'autonomie dans la gestion du routage multipoint entre domaines. Les protocoles actuels n'ont pas été conçus pour des systèmes autonomes multiples et ne peuvent restreindre la diffusion de l'information de routage selon les politiques et règles que les administrateurs pourraient désirer utiliser. La croissance du multipoint IP est fortement limitée si tous les routeurs doivent comptabiliser l'information de routage sur la totalité du réseau. La seule façon de dissimuler l'information est d'opter pour une topologie de routage hiérarchique. Pour le routage intra-domaine, deux normes — une pour le mode dense et une pour le mode clairsemé — seront sans doute nécessaires si l'on veut tenir compte de la grande variété d'applications multipoint.

L'intégration des réseaux satellites à l'Internet est en voie de réalisation. S'appuyant sur les infrastructures du câble et des communications par satellite, les fournisseurs de services Internet construisent de nouveaux modèles d'affaire et lancent des services à valeur ajoutée qui incluent le multipoint IP. Les protocoles courants de routage et de multipoint pour l'Internet ont été conçus pour des performances optimales en supposant des liaisons de communication symétriques et bidirectionnelles. Les solutions actuelles mises d'avant, telle l'utilisation d'un tunnel pour le canal de retour, sont non-optimales et ne devraient être utilisées qu'à court terme.

Signification pour les communications tactiques

Aucun des protocoles de routage multipoint en existence aujourd'hui n'opérera vraisemblablement de façon satisfaisante dans un environnement militaire tactique. Pas un n'a été spécifiquement conçu pour fonctionner dans des environnements sans fil et à bande étroite si caractéristiques des réseaux militaires, là où la symétrie des liaisons est parfois rompue, bien qu'étant habituellement requise pour une opération normale. De plus, la plupart de ces protocoles ont des caractéristiques de fiabilité, d'adaptabilité et de résistance insuffisantes qui devront être améliorées.

Suggestions de travaux de recherche

Quelques-unes des défaillances qui demandent une attention particulière, incluent:

i. la capacité d'acheminer le trafic vers des destinations choisies selon des politiques bien définies et pour lesquelles des niveaux de service (QoS) ont été sanctionnés et garantis;

ii. la capacité de limiter la croissance du tas dinformation de routage tout en permettant au routage multipoint de fonctionner selon les politiques et QoS désirées à l'intérieur d'un amalgame de systèmes autonomes;

iii. la capacité de limiter la croissance du tas dinformation de routage tout en permettant au routage multipoint de fonctionner selon les politiques et QoS désirées à l'intérieur même d'un système autonome partitionné;

iv. la capacité de s'adapter à un vaste ensemble de réseaux hétérogènes, parmi lesquels on retrouve des taux de transmission à faible débit et des liaisons unidirectionnelles.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

Multicast routing is a relatively new technological development, which is almost still experimental. During the past few years, one has witnessed the emergence of several multicast protocols, which are competing to provide efficient mechanisms to deliver IP traffic to user groups scattered throughout the Internet. Even though the development of the Internet proceeds by a succession of appendages instead of following a well-defined plan, the multiplicity of experimental protocols and the absence of any well-established standard protocol for multicasting indicate that multicast routing has many solutions and that no one implementation can provide the most satisfactory characteristics in every situation. An analysis of multicast routing architectures by Ballardie and Crowcroft [1] concludes that there are trade-offs to consider for each of the different methods: each method has its place in the range of multicast solutions, just as each of the unicast routing protocols has its place in the Internet. This is partly due to the diversity of multicast applications and correspondingly, to the wide variety of multicast application requirements. Examples of these applications include distribution of software updates, propagation of realtime data, efficient network news delivery, distance learning classes, video conferences and distributed interactive simulation (DIS). The latter, in particular, has strict requirements in terms of join latency, group membership dynamics, group sender populations, far exceeding the requirements of many other multicast applications. This paper will show that much work is still needed to advance the state of the multicast technology.

An exhaustive survey of the many multicast protocols that are currently proposed in the open literature is beyond the scope of this report. Instead, the main deficiencies and the challenging design issues of multicast protocols will be illustrated by focusing on a few of the most popular multicast protocols being developed on or experimented with today by the Internet Engineering Task Force (IETF).

One should note that the work discussed herein concentrates on network layer multicast. There is no discussion of transport layer (reliable) multicasting, which is a different problem space involving end-to-end delivery.

2.0 Background on Unicast Routing in the Internet

IP multicasting faces many of the same issues as the unicast routing protocols now used in the Internet. In fact, multicasting comes at a time when the "normal" routing infrastructure has not fully stabilized as evidenced by the many non-standard routing protocols currently being used in the Internet, especially with regard to the interconnection between organizations' networks and service providers. In this respect, the "commercialization" of the Internet creates routing control requirements that exceed the policy routing capability offered today by unicast protocols. Even more stringent requirements are present when multicasting is involved. For instance, high performance applications like video conferences (local, regional, national, international), distance learning classes and distributed simulation may, in many cases, require well coordinated multicast policy and superior access-control, management and QoS support. Furthermore, unicast and multicast protocols interact in several ways: they are not independent. For example, most multicast protocols interoperate with BGP-4 (a Border Gateway unicast Protocol [8]) to ensure interdomain connectivity. Many multicast protocols that are independent of the underlying unicast algorithm — e.g. PIM (Protocol Independent Multicast), CBT (Core Based Trees) — are forced to follow the policies specified by unicast routing. Because it affects the performance and design of the IP multicast infrastructure, the state of the unicast routing technology is summarized in this chapter.

2.1 Unicast Routing Algorithms

The organization of unicast routing — the structure that glues together routers of the worldwide Internet — consists of three basic routing algorithms. This characteristic can be used to group all unicast routing protocols into three distinct families:

1. *Distance Vector Protocols*, also referred to as "Bellman-Ford" protocols, are based on a distributed version of a very simple shortest-path computation algorithm. The algorithm's complexity is $O(MN^2)$, where N is the number of nodes and M *is* the number of links in the network. Each node keeps the distances separating it from the other destinations in its routing table.
2. *Link State Protocols* are based on the Shortest Path First (SPF) algorithm developed by E.W. Dijkstra. SPF converges in $O(MlogM)$ iterations while Bellman-Ford converges in $O(NM)$, where N is the number of nodes, which is generally of the same order of magnitude as the number of links M. For large networks with many links, the use of one algorithm over the other can make a sizable difference. Link state protocols are based on the "distributed map" concept: all nodes maintain a copy of the network topology in their routing database.
3. *Path Vector Protocols* are based on a loop-protection algorithm in which each routing update carries the full list of transit networks, or autonomous systems, traversed between the source and the destination nodes. So, path vector protocols are significantly different than distance vector protocols: rather than maintaining just the cost to each destination, each router keeps track of the exact path used.

Most specialists favour link state protocols over the distance vector variety for the following reasons:

- fast convergence;
- loopless convergence;
- support of precise metrics and, if needed, multiple metrics;
- support of multiple paths to a destination; and,
- separate representation of external routes.

2.2 Unicast Routing Protocols

The Internet can be described as a loose interconnection of networks belonging to many owners. One usually distinguishes three levels of networks:

1. organizations such as companies and institutions generally manage an internal network;
2. most organizations' networks are connected to the Internet through a "regional" provider which manages a set of links covering a state, a region, or maybe a small country; and,
3. a "transit" provider that ensures worldwide connectedness.

From a routing point of view, the Internet is split into a set of domains, also called autonomous systems (AS), i.e. a set of routers and networks under the same administration, usually organizational, regional or transit as mentioned above. This division into domains provides an hierarchical structure which permits better management of the routing overhead and the size of the routing tables as well as policy routing along the traffic path. In order to support each domain's autonomy and heterogeneity, routing consists of two distinct components: intra-domain (interior) routing, and inter-domain (exterior) routing. Intra-domain routing provides support for communication between hosts where datagrams traverse transmission and switching facilities within a single domain. Inter-domain routing provides similar support between domains. Border routers (gateways) are entry points in adjacent domains that forward packets across domain boundaries. The entities responsible for exchanging inter-domain routing information are sometimes called route servers and are usually collocated with the border routers. This role is achieved by exchanging "reachability information" between adjacent domains through an exterior routing protocol.

Current work on intra-domain routing within the Internet community has converged on the development of one standard interior gateway protocol for IP networks: OSPF (Open Shortest Path First). On the other hand, work on inter-domain routing has diverged in two directions: one is best represented by the *Border Gateway Protocol/Inter-Domain Routing Protocol* (BGP/IDRP) architectures and another is best represented by the *Inter-Domain Policy Routing* (IDPR) architecture. The two architectures are quite complementary and should not be considered mutually exclusive.

Most popular protocols of the unicast routing technology are summarized in Table 1; a brief description of the protocols follows.

TABLE 1. Select Group of Unicast Routing Protocols

Protocol Hierarchy	Most popular unicast protocols in today's Internet		Status[1]	Protocol Family
Interior	RIP	Routing Information Protocol [2]	Hist	Distance Vector
	OSPF	Open Shortest Path First [3]	Std	Link State
	IGRP/ EIGRP	Interior Gateway Routing Protocol/ Enhanced IGRP [4][5]	n/a[2]	Distance Vector
	IS-IS	Intra-Domain Intermediate System to Intermediate System Routeing Protocol [6][7]	Info	Link State
Exterior	BGP-4	Border Gateway Protocol [8]	D-Std	Path Vector
	IDRP	Inter-Domain Routing Protocol [9]	n/a[3]	Path Vector
	IDPR	Inter-Domain Policy Routing [10]	P-Std	Link State
	SDRP	Source Demand Routing Protocol [11]	Info	"Virtual link" by Encapsulation
	Tunnel	IP Encapsulation within IP [12]	P-Std	"Virtual link" by Encapsulation

[1] Classification of the protocol as per RFC 2400, September 1998 [13];
Std=Standard, D-Std=Draft-Standard, P-Std=Proposed-Standard, Exp= Experimental, Info= Informational, Hist=Historic, I-D= Internet-Draft not on the Standards track

[2] Proprietary protocol defined by Cisco Systems Inc.

[3] A protocol of the OSI routing framework. See also RFC 1745 and 1863.

2.2.1 RIP

Routing Information Protocol (RIP) is used widely in the Internet. However, being the simplest but also the oldest of the protocols listed in Table 1, RIP is plagued with severe technical limitations. The Internet Architecture Board (IAB) urges the Internet community to implement OSPF2 as the default interior gateway protocol for IP networks.

2.2.2 OSPF

As of April 1998, OSPF Version 2 is the official Interior Routing Protocol (open standard STD 54 [3]) for the Internet. The only real contender to OSPF today is the Enhanced Interior Gateway Routing Protocol (EIGRP). OSPF is based on link-state algorithms that permit rapid route calculation with a minimum of routing protocol traffic. In addition to efficient route calculation, OSPF supports hierarchical routing, load balancing, and the import of external routing information. The recommended maximum size for an OSPF area is 200 routers [58].

2.2.3 IGRP/EIGRP

Interior Gateway Routing Protocol (IGRP) and Enhanced Interior Gateway Routing Protocol (EIGRP) are not Internet standards, rather these are proprietary protocols defined by the network equipment manufacturer, Cisco Systems Inc. IGRP was developed before the IETF defined a new standard to replace RIP. IGRP and EIGRP include corrections for the known deficiencies of distance vector protocols like RIP. Some of the improvements include composite metrics, conservative protection against loops, multipath routing, and handling of default routes. Key elements of their operation have been patented by Cisco. This improved distance vector technology can perhaps compete with the reliability of the link state technology only by becoming equally complex. From a technical point of view, most experts believe that link state protocols are "better", but the strong minority of experts at Cisco does not accept this conclusion. From a user's point of view, and given that EIGRP and OSPF are both offering very acceptable performance, most are likely to insist on an "open standard" protocol such as OSPF to maintain "vendor independence". Choosing OSPF means that one can buy products from several vendors and benefit from the competition.

2.2.4 IS-IS

The Intermediate System to Intermediate System (IS-IS) protocol is part of the Open Systems Interconnection (OSI) routing framework. The IS-IS protocol was designed for use with ISO's connectionless network layer protocol, CLNP. Today, IS-IS has been modified to handle other protocols as well, most notably IP. In fact, the differences in quality and performance between IS-IS and OSPF are not very important: many of the new ideas developed in IS-IS were later adopted by OSPF. Further, there are fewer supporters of IS-IS than there are for OSPF. The deployment of IS-IS is very limited: it is used in some digital cellular systems such as Cellular Digital Packet Data (CDPD) and in some Internet backbones (including the old NSFNET backbone). Novell NetWare uses a minor variant of IS-IS (NetWare Link Services Protocol or NLSP) for routing IPX (Internet Packet Exchange) packets.

2.2.5 BGP-4

Between ASs, the recommended routing protocol on the Internet is the Border Gateway Protocol (BGP), Version 4. BGP is not yet a fully-approved Internet Standard (it is at the Draft-Standard level) although it is widely used. The protocol runs over TCP (Transmission Control Protocol) and has been criticized by some for its sensitivity to network congestion. Like other exterior gateway protocols, BGP has been designed to allow many kinds of routing policies to be enforced in the inter-AS traffic. Typical policies involve political, security, or economic considerations. Policies are manually configured into each BGP router; they are not part of the protocol itself. Weights can be assigned to some ASs in order to assert preferences and to represent policy constraints. BGP-4 supports Classless Inter-Domain Routing (CIDR) which allows routing table aggregation (BGP-4 was deployed in 1994, just in time to avoid the collapse of the Internet from the explosion in the size of its routing tables). Certain functionality of BGP-4 borrows heavily from IDRP, which is the OSI counterpart of BGP. The Inter-Domain Routing working group of the IETF is chartered to plan a smooth transition from BGP-4 to IDRP to support forwarding of IP datagrams across multiple ASs. IDRP is seen by the working group as a protocol that will support IPv4 as well as the next generation of IP (IPv6).

2.2.6 IDRP

The Inter-Domain Routing Protocol (IDRP) is part of the OSI routing framework. The deployment of IDRP is very limited and probably non-existent outside experimental circles. According to Huitema, IDRP is so similar to BGP-4 that at one time there was a consensus within the IETF working groups not to develop a version 5 of BGP, but rather simply to use IDRP. IDRP includes several enhancements to BGP-4, including the support of multiple-addressing families and variable address lengths, or the organization of AS into "confederations", which is used to aggregate the AS path information.

2.2.7 IDPR

Inter-Domain Policy Routing (IDPR) was developed by another working group[1] of the IETF at the same time that BGP was designed. Since July 1993, IDPR has been a "proposed standard" for the Internet, but there is no evidence of any large-scale deployment, let alone usage outside experimental circles. The common objective of IDPR and BGP is to provide aggregation at a higher granularity than the AS. IDPR does it using the powerful, yet complex, link state routing technology whereas BGP uses the path vector approach. IDPR supports the notion of "policy gateways", i.e., a set of border routers that interconnects, virtually, two or more ASs together.

2.2.8 SDRP and Tunnelling

The Source Demand Routing Protocol (SDRP) aims at defining a special case of routing where, on rare occasions, the packets must follow a specific sequence of relays (nodes) through the network to implement an particular policy. Tunnelling aims at achieving the same goal, but it is limited to a single "virtual" link. These source-initiated routing techniques rely on the same route selection mechanism (encapsulation of IP in IP) to satisfy a particular quality of service or accommodate any form of routing that is influenced by factors other than merely picking the shortest path. A source-demand routing architecture, used as the only means of inter-domain routing, has scaling problems because it does not lend itself to general hierarchical clustering and aggregation of routing and forwarding information. Currently, SDRP is not used outside experimental circles and tunnelling is

[1] The Inter-Domain Policy Routing Group of the IETF has concluded its activities shortly after developing a prototype implementation of IDPR.

still extremely limited, for example, to select an appropriate provider and to support multicasting in the Multicast Backbone (MBONE).

3.0 Multicast Routing Architectures

IP Multicasting on a single broadcast network is simple. Deploying multicast routing algorithms in a very large network is however, a complex task. So far there has been few attempts by the IETF to motivate a strategy for evolving the multicast routing development towards a specific target multicast architecture. One such vision was presented by the Inter-Domain Multicast Routing (IDMR) working group in an Internet-Draft entitled "*Hierarchical Multicast: Architecture & Transition Strategy*". The document presented a framework to converge the multicast routing infrastructure on the unicast routing infrastructure. The draft silently expired in December 1996 without being renewed or replaced. A more recent attempt is the Border Gateway Multicast Protocol/Multicast Address-Set-Claim (BGMP/MASC) architecture, also developed by the IDMR working group. BGMP/MASC provides mechanisms to realize inter-domain multicast on a global scale in the Internet. This architecture allows existing protocols to operate autonomously within each domain. In a way, these efforts are more conciliative than anticipative proposals, i.e., better at unifying than leading development of new architectures as one can appreciate when looking at the diversity of the multicast protocols currently being proposed.

3.1 Properties of Multicast Routing Protocols

A list of the desirable properties of a multicast routing protocol is fairly long. Currently, there is no consensus on which properties should come first in the design of the multicast Internet architecture. However, much of the uniqueness found in the existing protocols depends on how each one of these protocols considers four closely related properties: *hierarchism*, *scalability*, *autonomy* and *Policy/QoS compliance*. Most support the first property, several the second but only a few support the third and fourth ones. It is not necessary for a single protocol to support all four properties to their fullest extent. However, it is desirable that the multicast routing infrastructure be capable to support all four properties if multicasting is to fulfil its role adequately in an Internet of ever increasing size and heterogeneity . Each of these properties is now examined in more detail.

1. *Hierarchism* — Hierarchical multicast routing is a way to hide routing information. Much like the unicast routing case, once a network reaches a certain size, the multicast routing overhead, the size of the multicast routing table and the frequency of the routing exchanges become so significant that some of the routers and the links become unstable. A hierarchical structure not only contains the routing information within an affected "region", but also better facilitates route aggregation to permit more scalable growth.

   At the beginning, the Internet's multicast infrastructure, mainly composed of the MBONE, was a flat, tunnelled, non-hierarchical topology which limited the ability to aggregate routing information and contain topological changes or operational problems that affected it. From a global perspective, this architecture is slowly evolving to one that has two hierarchical levels: inter- and intra-multicast regions. Much like the hierarchy present in unicast routing, each region independently chooses to run whichever multicast routing protocol that best suits its needs, and the regions interconnect via the "backbone region", which currently runs DVMRP. This popular architecture limits the changes required to support hierarchical routing to the border routers operating on the edge of the multicast region. The hierarchy is preserved when tunnels only exist within a region but are not allowed to "pass through" a border router (i.e. tunnels may terminate at a border providing both end-points of the tunnel lie within the same region).

Such a simple approach is not without its problems. DVMRP, a distance vector protocol, is widely considered inadequate for rapidly changing network topologies partly because routing information propagates too slowly. Its inability to detect routing loops and oscillating links is one of its main deficiencies. Until recently there were few alternatives. MOSPF, for instance, only provides routing facilities within an autonomous system of limited size; the emphasis is on efficient route computation. It does not include any provision for setting up tunnels—the reasoning is that tunnels are only a transition tool and that very soon the majority of the area's routers will be multicast-capable. Others like PIM and CBT simply ignore the problem by relying on an underlying unicast routing protocol. Things are however changing. There are now hierarchical versions of DVMRP, PIM and CBT, i.e., hierarchical DVMRP, HPIM (Hierarchical PIM) and OCBT/CGBT (Ordered CBT/Core Group Based Trees) respectively. The multicast version of BGP (BGMP) also supports hierarchical (multicast) routing by segmenting the address space so that each segment represents a different hierarchical level. This scheme not only "collapses" routing information, and contains the routing problems, but also can prevent multicast packets from travelling beyond a particular hierarchy level — a very desirable characteristic.

2. *Scalability* — Much like hierarchism, controlling the amount of group state information maintained in the network, the bandwidth consumption (link utilization) and processing costs leads to a scalable multicast protocol. A multicast protocol that scales well is a protocol that is resource-efficient and maintains good performance regardless of the distribution of the multicast group members throughout the network. It would be difficult, and too complex, for a protocol to dynamically adjust its routing parameters to every small change in the distribution of the group members. Instead, IP multicast routing algorithms and protocols generally follow one of two coarse approaches.

   The first approach is based on the assumption that the multicast group members are densely distributed throughout the network and bandwidth is plentiful, i.e., almost all hosts on the network belong to the group. So-called "dense-mode" (-DM) multicast routing protocols include DVMRP, MOSPF and PIM-DM. The second approach is based on the assumption that the multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available, e.g., across many regions of the Internet. It is important to note that sparse-mode does not imply that the group has a few members, just that they are widely dispersed. "Sparse-mode" (-SM) routing protocols include CBT and PIM-SM. As one can see, dense and sparse are just two extreme situations; one could imagine groups which are semi-sparse, medium-dense or anything else in between.

   In each case, the resources are controlled by setting up at start-up opposed mechanisms for reaching the multicast group members. The default forwarding action of the dense-mode multicast routing protocols is to forward traffic, while the default action of a sparse-mode multicast routing protocol is to block traffic unless it is explicitly requested.

3. *Autonomy* — A major technical hurdle to the deployment of multicast applications throughout the Internet today is the lack of a standard protocol for inter-domain (exterior) multicast routing. PIM, MOSPF and DVMRP, the interior/wide-area multicast routing protocols in common use, are not designed for multiple autonomous systems that do not necessarily want to share all their routing information. They blindly forward all routing information to all known routers. Growth is severely limited if all routers have to contain all multicast routing information for the whole Internet.

   The IDMR working group of the IETF has recently produced a draft of BGMP which is a multicast version of the inter-domain Border Gateway Protocol. However, BGMP is still under research at this point. It is not expected that the protocol will see widespread deployment very soon.

4. *Policy/QoS Compliance* — In its broadest sense, Policy/QoS routing refers to any routing that is influenced by factors other than merely picking the shortest-path as the preferred route, such as finding a path that provides a particular quality of service. A multicast protocol that considers QoS in its routing phase can create a tree better suited to the needs of QoS-sensitive applications. Existing multicast protocols are constrained to only a single path but QoS introduces a mechanism to provide multiple routes between source and destination. The requirement for Policy/QoS routing also appears with the commercialization and marketability of the Internet (e.g. selecting a service provider because of a guaranteed quality of service).

   Currently, services over the Internet are limited by the best-effort nature of the network. Traditional Internet routing protocols do not consider QoS metrics. Therefore, it is not surprising that very few multicast routing protocols were designed to be Policy/QoS compliant. The new protocols that are being proposed partially address these issues. For instance, QoSMIC (QoS Multicast Internet protoCol [31]), recently being released as Internet-Draft by the University of Toronto, includes the main concepts of the YAM protocol and introduces several new ideas that make it flexible but also quite complex. PTMR (Policy Tree Multicast Routing), a development sponsored by Cisco Systems Inc., is based on PIM-SM and aims at attaining policy-sensitive data packet delivery in an Internet-wide multicast. The difficulty is that Policy/QoS is fundamentally an end-to-end issue which involves many components of the network resources. The Internet infrastructure evolved rapidly but without fully integrating the Policy/QoS routing concepts that were explored during its development. The current trend is to propose new network architectures (e.g. Asynchronous Tranfer Mode (ATM), IPv6) where such support can more easily be integrated.

3.2 Multicast Routing Trees

Multicast routing protocols build routing trees for the dissemination of messages to a select group of other stations. The types of multicast trees built by the routing algorithms can be roughly divided into two families:

1. *Source Based Tree*, also referred to as "*Shortest Path Tree*", is a tree where the receiving group is assumed to be *fairly dense* and the sender initiates the multicast assuming all routers in the network are interested in receiving the multicast. The routing tree is formed

along the shortest path between each sender and receiver, hence the overhead at a router is O(*NS)*, where *N* is the number of multicast groups and *S* is the number of sources in the group.

2. *Group Based Tree*, also referred to as "*Group-Shared Tree*" or simply "*Shared Tree*", is a single (shared) tree created for all senders and receivers in the group. The receiving group is assumed to be *fairly sparse* so receivers initiate their own connection to the tree. The router does not have to maintain information about each source in each group, but has instead a single entry for each group. The overhead at a router is O(*N)*, where *N* is the number of multicast groups. This gives the shared tree approach superior scalability. However, because each packet no longer travels over its shortest path to each receiver, shared trees incur longer average delay in the delivery of a data packet. Wall [14] has proven that the maximum delay bound of an *optimal centre based tree*[1] is twice that of a shortest-path tree.

[1] A shared tree with a node optimally positioned to act as a meeting point between a sender and group receivers.

4.0 Overview of Multicast Routing Protocols

In this section the state of the multicast routing technology for the Internet is presented. Some of the most popular multicast routing protocols are listed in Table 2. The coupling between these protocols and the unicast routing protocols is summarized in Table 3. The multicast protocols are briefly described thereafter.

TABLE 2. Select Group of Multicast Routing Protocols

Protocol Hierarchy	Most popular multicast protocols in today's Internet		Status[1]	Delivery Tree Family	Discovery Method
Interior (Dense Group Distribution)	DVMRP	Distance Vector Multicast Routing Protocol [15]	Exp	Source Based	Broadcast and Prune
	MOSPF	Multicast OSPF Protocol [16]	P-Std	Source Based	Explicit Join
	PIM-DM	Protocol Independent Multicast (Version 2) - Dense Mode [17]	I-D	Source Based	Broadcast and Prune
Interior or Wide-Area (Sparse Group Distribution)	PIM-SM	Protocol Independent Multicast - Sparse Mode [18]	Exp	Group Shared[2], Unidirectional	Explicit Join
	CBT	Core Based Trees Protocol Version 2 [19]	Exp	Group Shared, Bidirectional	Explicit Join
Exterior	BGMP	Border Gateway Multicast Protocol [20]	I-D	Group Shared, Bidirectional	Explicit Join

[1] Classification of the protocol as per RFC 2400, September 1998 [13]; Std=Standard, D-Std=Draft-Standard, P-Std=Proposed-Standard, Exp= Experimental, Info= Informational, Hist=Historic, I-D= Internet-Draft not on the Standards track

[2] PIM-SM is a hybrid protocol. A PIM-SM router has the option of switching to the source's shortest-path tree as soon as it starts receiving datagrams from the source station.

TABLE 3. Coupling between Unicast and Multicast Routing Protocols

Protocol	Protocol Family	Underlying Unicast Routing Requirement
DVMRP	Distance Vector	None (uses built-in RIP-like routing protocol)
MOSPF	Link State	OSPF
PIM-DM, PIM-SM, CBT	Independent	Any unicast routing protocol
BGMP	Path Vector	BGP4 + Multicast RIB (MBGP)

4.1 DVMRP

If there is one popular multicast routing protocol in use in the Internet, it must be DVMRP. Since 1992, it has been the central component of the MBONE, the first major experimental multicast routing network. It is used widely in the research community to transmit the proceedings of various conferences and to permit desktop conferencing.

DVMRP is a very simple distance vector routing protocol, quite similar to RIP. The major difference between RIP and DVMRP is that RIP is concerned with calculating the next hop to a destination, while DVMRP is concerned with computing the previous hop back to a source. Current implementations (mrouted Version 3.8 or higher) have extended DVMRP to employ the Reverse Path Forwarding (RPF) and Prune algorithm which is an improvement over the Truncated Reverse Path Broadcasting (TRPB) algorithm defined in the original RFC 1075 specification. In fact, DVMRP as specified in RFC 1075 is about to be declared a historic protocol [21]. The IETF is working on a third version of DVMRP which should reflect the implementations most widely used today throughout the Internet [22].

In addition to the RIP-like functions, DVMRP is designed to traverse networks that do not support multicasting. This is accomplished by manually setting up tunnels using three parameters: the destination router IP address, a metric that specifies the cost (essentially, a hop count) to use when computing the DVMRP distances, and a time to live (TTL) threshold that limits the scope of a multicast transmission. Since DVMRP is not very precise, it is difficult to choose the proper value for a tunnel's costs and thresholds. It is even difficult to choose the proper places to place tunnels: random connections may lead to surprising results.

DVMRP, like RIP, is plagued with the same severe technical limitations. For rapidly changing network topologies or group distributions, the routing information propagates too slowly. These limitations are exacerbated by the fact that early implementations of DVMRP did not implement pruning.

4.1.1 Hierarchical DVMRP

DVMRP was designed as an interior gateway protocol suitable for use within an autonomous system, but not between different autonomous systems. However, because of its tunnelling capability, DVMRP can manually interconnect ASs. As the number of subnetworks relying on DVMRP continues to increase, the size of the routing tables and of the periodic update messages will continue to grow. If nothing is done about these issues, the processing and memory capabilities of the DVMRP routers will eventually be depleted and routing will fail. To overcome these potential threats, a hierarchical version of the DVMRP has been proposed [23] at a meeting of the Association for Computing Machinery (ACM) in 1995. Hierarchical DVMRP proposes the creation of non-intersecting regions where each region has a unique Region-Id. The routers internal to a region execute any multicast routing protocols such as DVMRP, MOSPF, PIM, or CBT as a "Level 1" protocol. Each region is required to have at least one "boundary router" that is responsible for providing inter-regional connectivity. The boundary routers execute DVMRP as a "Level 2" protocol to forward encapsulated traffic between regions. The design accommodates the eventual addition of more levels of hierarchy and the use of protocols other than DVMRP at any level. To this day, the proposal has not been formally submitted to the IETF and, given the new developments in multicast routing technology, Hierarchical

DVMRP is not expected to be rushed to the Standards track of the IETF, despite its intrinsic simplicity.

4.2 MOSPF

MOSPF is a proposed standard of the IETF. At the moment, it is the only multicast routing protocol other than IGMP to reach the Internet Standards track. The MOSPF specification was published in March 1994, and several router vendors (e.g. 3Com, Proteon, Cisco) have implemented it.

MOSPF is a set of extensions built on top of the unicast OSPF Version 2 routing protocol. For this reason, it can only provide routing facilities within an autonomous system of limited size. MOSPF, unlike DVMRP, does not provide support for tunnels. To facilitate inter-AS multicasting routing, selected router are configured as “inter-AS multicast forwarders” and execute an inter-AS multicast routing protocol (such as DVMRP), which forwards multicast datagrams in a reverse path forwarding (RPF) manner.

With the link-state mapping capability of OSPF, MOSPF routers maintain a current image of the multicast tree topology. The intra-area trees are detailed and precise whereas in the case of inter-area routing, it is possible that incomplete trees are created because detailed topological and group membership information for each OSPF area is not distributed between OSPF areas. To overcome these limitations, topological estimates are made using information provided in summary-links advertisements originated by the source subnetworks.

MOSPF performs one shortest-path computation per combination of source and group. Since there are potentially as many sources as hosts in an area, and that the number of groups itself is likely to grow with the size of the AS, the number of computations that follow any routing update is likely to grow as the square of the size of the area. As the cost of each computation is of the order of $O(NlogN)$, there is a potential for saturating even the most powerful router’s CPU. To alleviate the problem, MOSPF routers do the computation “on demand” when they receive the first datagram of a group transmission. That is, for a given multicast datagram, all routers within an OSPF area calculate “in memory” the same source-rooted shortest path delivery tree when a router receives the first multicast datagram for a particular source-group pair. The information in memory is not aged or periodically refreshed, rather it is maintained as long as there are system resources available or until the topology (distribution of group-memberships) change.

Unlike DVMRP, where data packets are periodically flooded by routers not on the multicast tree, in MOSPF the link-state packets containing the state information for group membership are periodically flooded. The latter approach means that the first datagram of a group transmission does not have to be forwarded to all routers in the area.

MOSPF routers are required to eliminate all non-multicast OSPF routers when they build their delivery tree. This can create a number of potential problems when forwarding multicast traffic, including:

- multicast datagrams may be forwarded along sub-optimal routes since the shortest path between two points may require travelling through a non-multicast OSPF router;
- even though there is unicast connectivity to a destination, there may not be multicast connectivity;
- the forwarding of multicast and unicast datagrams between two points may follow entirely different paths through the internetwork.

Unlike unicast OSPF, MOSPF does not support the concept of equal-cost multipath routing. "Tie-breakers" have been defined to guarantee that should several equal-cost paths exist, all routers will agree on a single path through the area. However, MOSPF is the only multicast routing protocol that currently offers explicit support for multiple types of service (TOS). IP datagrams can be labelled with any one of five TOS, namely: minimum delay, maximum throughput, maximum reliability, minimum monetary cost, and normal service. MOSPF calculates a separate path for each {source, destination, TOS} tuple, using Dijkstra's algorithm.

4.3 PIM - Dense Mode

PIM-DM Version 2 has been under development since August 1998 by the newly formed Protocol Independent Multicast (PIM) working group of the IETF. The group is chartered to standardize and promote PIM (-DM and -SM) and to act as a consultant to any alternative proposals. An obvious competitor to PIM-DM today is MOSPF which is being promoted by an other group of the IETF, the MOSPF working group. Interestingly, the original specifications for PIM and MOSPF were concurrently developed nearly five years ago. Unlike MOSPF, PIM (-DM and -SM) is currently supported by only a few router vendors (mainly Cisco, Lucent Technologies).

PIM-DM is neither a distance vector nor a link state multicast routing protocol. It does not mandate the computation of specific routing tables: it simply supposes that such tables exist. Compared with multicast routing protocols with built-in topology discovery mechanisms (e.g. DVMRP with its own RIP-like unicast routing protocol, or MOSPF with its dependence on the information contained in the OSPF link-state database), PIM-DM has a simplified design, and is not hard-wired into a specific type of topology discovery protocol. However, such simplification does incur more overhead and generates traffic on some links that could be avoided if sufficient topology information is available.

In PIM-DM, the multicast routing is performed with a very simple algorithm: RPF and prune. The unicast routing table is used to determine whether a neighbour is upstream with each multicast source. If so, a multicast datagram is flooded on every other (downstream) interfaces until explicit prune messages are received. PIM-DM is therefore characterized by the periodic transmissions of broadcast and prune messages throughout the entire network. "Graft" messages are also used to re-establish the previously pruned branch on the delivery tree when group members appear on a pruned branch. This whole approach is similar to the one used by DVMRP except that it is independent of the mechanisms of a specific unicast routing protocol.

PIM-DM assumes that the point-to-point routes are symmetric (the path characteristics are the same in both directions). Packet duplication may occur when this is not the case causing routers to receive duplicate packets from the source along different paths. Duplicate datagrams can also occur when there are parallel paths to a source, particularly, if two routers have equal cost paths to a source and are connected on a common multi-access network. PIM-DM will detect such a situation and will not let it persist.

4.4 PIM - Sparse Mode

PIM-SM is currently an experimental Internet protocol. One of the objectives set by the IETF's PIM working group is the submission of PIM Version 2 (-DM and -SM) specifications to the Internet Engineering Steering Group (IESG) by April 1999 for consideration as an Internet Draft-Standard.

PIM-SM, like PIM-DM, is not dependent on any particular unicast routing protocol. Furthermore, PIM-SM control message processing and data packet forwarding are integrated with PIM-DM operation so that a single router can run different modes for different groups. Routing algorithm independence is a "double-edged sword": it simplifies multicast routing across heterogeneous domain boundaries, and it allows for the independent evolution of both unicast and multicast algorithms, but multicast routing is forced to follow the policies specified for unicast routing rather than its own separate routing policies. Further, implementation is made more complex in such cases.

PIM-SM is designed to address the potential scaling problems of the dense-mode multicast algorithms in large wide-area. It operates in each domain, which in this context means a contiguous set of routers that all implement PIM-SM and operate within a common boundary defined by PIM Multicast Border Routers.

PIM-SM allows group members to receive multicast data either over a shared tree, which receivers must first explicitly join, or over a shortest-path tree, which a receiver can create subsequently, in an attempt to improve delay characteristics between some active source, and itself. The change-over to a shortest-path tree may be triggered if the data rate from the source station exceeds a pre-defined threshold. Other criteria are possible but none have been defined at this time. When a receiver creates a shortest-path to a particular source, it prunes itself off the shared tree for that (source, group) pair, but will continue to receive data packets for the group over the shared tree from all other sources.

The shared tree is built around so-called rendezvous points (RPs), of which there may be several for robustness purposes. The initiator of each multicast group selects a primary RP and a small ordered set of alternative RPs, known as the RP-set. For each multicast group, there is only a single active RP.

The designers of PIM-SM wanted a receiver to have the choice of receiving data over a shared tree or a source-rooted tree. Here there is a trade-off between routers keeping less state on a shared tree and more state on a shortest-path tree. Also, as the number of shortest-path trees grow for a particular source, the amount of bandwidth consumed by the sum of the shortest-path trees increases overall compared to a single shared tree.

PIM-SM is considerably more complex than DVMRP or the MOSPF extensions. It requires routers to maintain a significant amount of state information to describe sources and groups. For example, Bootstrap messages [24] are periodically distributed within a domain to all routers to provide the location of the RPs. For this purpose, PIM-SM uses an algorithmic mapping (hash function) from multicast group address to RP and a hierarchical model to keep information about potential RPs as local as possible. Other examples, if there is more than one local router on a LAN, PIM-SM must elect a designated router (DR) to join at the RP or to encapsulate traffic to the RP. An assert mechanism is also required to choose a single preferred route to one upstream router since a RP-tree and a shortest-path tree for the same group may both cross the same multi-access network.

4.4.1 Hierarchical PIM

A hierarchical version of PIM (HPIM) has been proposed in [25] but has not yet reached the Internet-Draft level. The most important way in which HPIM differs from PIM-SM is that HPIM does not require advertisement of RPs to the senders and receivers of a group. Instead, each router in the multicast tree makes a local decision about where the next hop RP is based on the multicast address and a candidate RP list which is synchronized across all RPs in the same scope at the same level.

Multicast addresses are allocated in bands which determine the scope of the session in a similar way to administrative scopes are handled in DVMRP. Candidate RP routers — a PIM router that is capable of being a RP and is configured as being available to be an RP— are given a "level" in a global hierarchy. Synchronization between RPs is achieved by having the candidate RPs at level *n* receive each others candidate RP announcements, and build a candidate RP list, which they then distribute to the level *n-1* routers. Typically there will be a small number of levels in the hierarchy (e.g. 5 to 10) and a relatively small number of RP routers at each level in each scope area.

Unlike PIM-SM, the amount of state that must be held anywhere to maintain the candidate-RP lists in HPIM is independent of the number of multicast groups or the number of senders and receivers.

4.5 CBT

CBT Version 2 is an experimental protocol progressing through the Inter-Domain Multicast Routing (IDMR) working group of the IETF. The CBT multicast architecture was developed shortly before the emergence of PIM.

Some claim that PIM has adopted the "good parts" of CBT, and has augmented CBT's features in order to allay CBT's disadvantageous properties, the most prominent of which is the potential for sub-optimal paths (which usually equates to delay) between two receivers. However, as PIM currently stands, it is debatable whether the gain obtained in terms of performance and delay is considerable enough to justify the additional complexity.

CBT and PIM-SM share the same objectives. Both are designed to address the potential scaling problem of the source-based multicast algorithms. They are the first attempts at providing Internet-wide and inter-domain multicasting capability. Their degree of success achieved towards this goal so far is still a matter of debate.

Both CBT and PIM-SM are independent of the underlying unicast routing algorithm used. To establish paths between senders and receivers, they only need to access a separate multicast routing table. In both, a single shared tree is created for all senders and receivers in the group, and receivers initiate their own connection to the tree through a well known router, called the rendezvous point and the core in PIM and CBT respectively. Support for several active cores was provided in Version 1 of CBT, but because of loop problems this approach was abandoned in later versions.

The capability to change from a RP tree to a shortest-path tree is the main difference between PIM-SM and CBT. Additionally, CBT uses "hard states." Messages are explicitly acknowledged and repeated after a time-out. PIM-SM uses "soft-states." Join messages are repeated at regular intervals, the states are cached and simply "disappear" if the information is not refreshed.

It should be noted that the current specification of CBT is not backwards-compatible with either the original version (CBT Version 1) nor with the third version [26] which was released in March 1998 as an Internet-Draft. Version 1 was specified with multiple cores which induced several problems, such as routing loops during times of underlying unicast instability, and the failure to build a connected tree, even when the underlying routing was stable [28].

4.5.1 Hierarchical CBT

There are at least two hierarchical versions of the CBT protocol: the Ordered Core Based Tree (OCBT) and the Core Group Based Trees (CGBT). The former is discussed in Section 4.7.2 while a brief description of the latter follows.

CGBT was presented by Yuan-Chi Chang, a graduate student of the University of California, Berkeley, in March 1996 at a meeting of the IDMR working group. CGBT [27] is a modification of the CBT protocol with the capability of dynamically splitting and merging the CBTs in response to the distribution of participants' locations and the delay requirements of the application. Tree splitting occurs and a new core is formed when the size of the downstream distribution tree exceeds a certain tree metric. Tree metric tables are based on the delay requirement of multicast applications and the network topology information, including the average node degree. The designers claim improved delay performance over CBT and good scalability. There is currently no Internet-Draft for this multicast protocol.

4.6 BGMP

Until recently, efforts have been concentrated on extending single-domain techniques to wide-area networks. BGMP, as specified in a recent Internet-Draft by the IETF IDMR working group, provides mechanisms to realize inter-domain multicast on a global scale in the Internet. This is early work and many details remain unresolved.

BGMP builds bidirectional shared trees for active multicast groups, and allows receiver domains to build unidirectional source-specific, inter-domain, distribution branches where needed. Building upon concepts from CBT and PIM-SM, BGMP requires that each multicast group be associated with a single root. However, in BGMP the root is an entire exchange or domain, rather than a single router, and the root is therefore referred to as the root domain. BGMP assumes that at any point in time, different ranges of the class D space are associated with different domains. This is accomplished with the use of a complementary protocol like the Multicast Address-Set Claim (MASC) protocol [30]. MASC dynamically allocates multicast address ranges to domains from which groups initiated in the domain get their multicast addresses. Each of these domains then becomes the root of the shared domain-trees for all groups in its range. BGMP allows any existing multicast routing protocol to be used within individual domains. The set of addresses claimed and obtained by a domain are advertised in BGP4+.

BGMP uses TCP as its transport protocol. This eliminates the need to implement message fragmentation, retransmission, acknowledgement, and sequencing. The BGMP and BGP transport interfaces are distinct. Such an approach provides protocol independence and facilitates distinguishing between protocol packets. Two BGMP peers, run by the border routers of separate domains, establish a TCP connection between one another, and exchange Join/Prune Updates as group memberships change. BGMP does not require periodic refresh of individual entries. However, the BGMP protocol state is refreshed by "keep-alive" messages sent periodically over TCP.

4.7 Other Recent Proposals

4.7.1 QoSMIC

QoSMIC stands for Quality of Service sensitive Multicast Internet protoCol. The protocol is an Internet-Draft [31] of the IDMR working group and is authored by Anindo Banerjea and Michalis Faloutsos, both with the University of Toronto, and Rajesh Pankaj from QUALCOMM Inc. QoSMIC supports both shared and source-specific trees. In both trees, the destination is able to choose the most promising among several paths. QoSMIC starts with a shared-tree and switches to a source-specific tree when necessary i.e. to meet some QoS requirement or for load-balancing (active sources). Dynamic routing information is collected and used without relying on a link state exchange protocol to provide it. QoS metrics include end-to-end delay and packet loss ratio. This is unlike BGMP, PIM, CBT and MIP which have no QoS support. In QoSMIC, instead of a Rendezvous Point or core as in PIM-SM or CBT, there is a "managing router" which is in charge of tree construction. Data need not flow through it. This reduces the "core placement" problem to some extent, since poor placement of the managing router should simply affect join and leave latency and not actual data flow. In simulation, QoSMIC built very efficient trees.

4.7.2 Ordered CBT and HIP

Clay Shields and J. J. Garcia-Luna-Aceves, both from the University of California, Santa Cruz, showed that Version 1 of CBT could form loops during times of underlying unicast instability and that it could consistently fail to build a connected tree, even when the underlying routing is stable. Consequently, they designed the Ordered Core Based Tree (OCBT) protocol [28] to remedy these shortcomings of the early version of CBT protocol. OCBT is a hierarchical multicast protocol where every core maintains an integer logical level i.e. a label indicating the cores' place in the hierarchy of cores. OCBT limits control messages to within a particular logical level and distributes the processing of control messages over a larger number of cores. Besides being proven to be loop-free at all times, OCBT relies directly on the underlying unicast routing protocol, e.g. BGP. Further, in July 1998, Shields and Garcia-Luna-Aceves introduced a new protocol called the HIP protocol (HIP) [29] that uses OCBT as the inter-domain (wide-area) routing protocol in a hierarchy that can include any multicast routing protocol at the lowest level. While maintaining the same functionality of other hierarchical multicast schemes, HIP approaches the problem in a very different manner. Routing between the domains is based solely on the unicast routing tables: the lower-level domains do not need to be explicitly named and no separate routing needs to occur to build paths to domains. Data traffic flows over a single tree, and while higher-level control messages may be encapsulated for transmission across a region, data packets are never encapsulated or duplicated. The tree itself is always built with an attempt at forming the average shortest path to the centre point for the group. The HIP protocol introduces the idea of a virtual router (VR) that is formed by all border routers of a domain operating in concert to appear as a single router in the higher-level tree. There is currently no Internet-Draft for either OCBT or HIP.

4.7.3 PTMR

The Policy Tree Multicast Routing (PTMR) protocol [32] was developed by Horst Hodel, from the Swiss Federal Institute of Technology, during a sabbatical at Cisco Systems Inc. of San José, California. The protocol aims at attaining policy-sensitive data packet delivery in Internet-wide multicasts across various domains, even under asymmetric conditions. PTMR's characteristic feature is the forwarding of multicast packets in accordance with any underlying multicast-relevant routing, including policy routing (supporting source-specific policies as well as shortest-path and QoS criteria). PTMR is based on PIM-SM and applies receiver-initiated, source-originating tree construction. PIM-SM provides for the source/receiver handshake and for initial source specific trees before switching to PTMR mode. The PTMR-tree is formed by Policy Routes, i.e. macroscopic paths from source to group member Multicast Domains (MDs) given by a sequence of MDs which satisfy the policy requirements of both the source and the involved domains and supports the requested QoS. There is currently no Internet-Draft for PTMR.

4.7.4 MIP

Multicast Internet Protocol (MIP) [33] can construct both group-shared and shortest-path multicast trees and accommodates two interoperable modes of tree construction; namely, sender-initiated and receiver-initiated, which makes it flexible for use in a wide range of applications with different characteristics, group dynamics, and group sizes. Just as with PIM, MIP is independent of the underlying unicast routing and assumes that the link costs are symmetric. Instead of using the idea of "soft-state" to maintain multicast routing information, MIP uses "diffusing" computations to update and disseminate multicast routing information. Under stable network conditions, MIP has no control message overhead to maintain multicast routing information.

4.7.5 DRP

The Designated Rendezvous Point (DRP) protocol [34] is based on a hybrid of PIM and CBT. The main difference between DRP and PIM or CBT is that DRP dynamically elects a router within a multicast region to perform the Rendezvous Point or Core functions. This centralized server is called the Designated Rendezvous Point. The DRP intelligently selects RPs for the multicast groups. If a session has special QoS requirements, DRP can select a RP in such a way that the constructed RP-tree conforms to them. A hierarchical architecture along with multicast address scope classification is proposed as future work.

5.0 Comparison

This section offers various comparisons among the multicast protocols listed in Table 2. In as much as it is possible, and to avoid inappropriate comparisons, only protocols belonging to a same or similar hierarchy, i.e., Interior Dense-group, Interior Sparse-Group or Exterior, are compared. To help see the trade-offs involved when choosing a multicast protocol, the section begins with a general comparison of the types of multicast trees constructed by the protocols.

5.1 Shared Trees vs Shortest-Path Trees

Shared trees scale more favourably than source-rooted trees, but have larger average delays. A group-based architecture provides a significant improvement in the overall scaling factor of the source-based tree architecture, from *SN* to just *N* (see Section 3.2). This is the result of having just one multicast tree per group as opposed to one tree per (source,group) pair.

The primary trade-off introduced by the shared tree architecture is a reduction in the overall amount of network states that must be maintained (given that a group has a significant proportion of active senders) and the potential increase in delay imposed by a shared delivery tree. Further, as the number of shortest-path trees increases for a particular source, more overall bandwidth is consumed by the sum of the shortest-path trees than by a single shared-tree.

A source-based architecture does not offer very favourable scaling characteristics for wide-area multicasting. For example, DVMRP and PIM-DM incur the pruning-state overhead on routers that are not on the multicast tree whereas group-based trees prevent data flow where it is not needed. For MOSPF, the overhead of Dijkstra computations does not scale to internetwork-wide multicasting.

Simulation results (see Section 5.3) show that group-based trees incur, on average, a 10% increase in delay over shortest-path trees. Even for real-time applications such as voice and video conferencing, a group-based tree may indeed be acceptable, especially if the branches of that tree are high-bandwidth links, such as fibre optics. However, the increased delay may not be acceptable if a portion of the delivery tree spans low bandwidth links.

A consequence of one shared delivery tree is that the cores of a particular group can potentially become traffic “hot-spots” or “bottlenecks”. This has been referred to as the traffic concentration effect.

Core (or Rendezvous Point) placement and management are other issues that may be seen as disadvantageous to the group-based approach. In particular, dynamic placement is a complex problem. It has been shown that dynamically changing the form of a multicast tree is usually not worthwhile for dynamic multicast groups in terms of efficiency benefits, since any benefit is likely to be short-lived. The selection of cores and their placement are important topics for further research. As finding the centre for a group is an NP-complete problem which requires knowledge of the whole topology, alternative practical forms are based on heuristic centre placement stategies.

Finally, it would seem that because of the nature of their forwarding mechanism, shortest-path tree schemes cannot achieve load balancing without the danger of loops forming in the multicast tree topology. Shared tree schemes have the ability to achieve load balancing, i.e., to forward incoming packets over different links.

5.2 Dense Group Distribution: DVMRP, MOSPF and PIM-DM

There does not appear to be any simulation study in the literature comparing the overall performance of DVMRP, MOSPF and PIM-DM operating under similar multicast environment. Table 4 summarizes some of the key points that were made in Sections 4.1 to 4.3. MOSPF is a very simple add-on feature for an OSPF network if the routers have enough CPU resources to perform the computations that it mandates. Besides, it is the only multicast routing protocol that supports QoS. Depending on the rate of development of the Internet multicast routing infrastructure, where support for QoS-sensitive application could become an important requirement, QoS support may play a major factor in promoting the use of MOSPF in the years to come. In the short term though, despite the technical superiority of link state technology used in MOSPF, it is likely that DVMRP will remain popular and PIM-DM will gain gradual acceptance because of their ties with the sparse group distribution protocols. DVMRP has grown in popularity through the MBONE experiment. Each region of the MBONE can choose to run whichever multicast routing protocol best suits its needs, but the regions interconnect via the "backbone region," which initially (and still does) ran DVMRP. Therefore, it follows that a region's border router must interoperate with DVMRP, a requirement that most multicast routing protocol designers try to fulfil in order to help promote the acceptance of their own protocols.

TABLE 4. Comparison of DVMRP, MOSPF and PIM-DM

Metric	Protocol		
	DVMRP	MOSPF	PIM-DM
Protocol Status	Experimental Internet Protocol	Proposed Internet Standard	Internet-Draft
Popularity	High. Selected by most to interoperate with sparse group distribution protocols like CBT, PIM-SM	Moderate	Emerging. Used to be a proprietary protocol defined by Cisco Systems Inc.
Routing Technology	Distance Vector. Reacts slowly to changes and prone to routing loops	Link State. Reacts quickly to changes, loop-free, but computationally intensive.	Routing independent. Relies on underlying unicast routing protocol capability
Handling of non-multicast capable routers	User manually sets up tunnels to reach multicast capable routers	Automatically eliminates all non-multicast OSPF routers from the SPF tree	Relies on underlying unicast routing protocol capability
Underlying unicast routing requirement	None	OSPF	Any unicast routing protocol
QoS support	1 metric through tunnelling	5 types of service	None. Assumes that point-to-point routes are symmetric
Hierarchism	Not supported although a hierarchical version of DVMRP has been discussed	2 Levels	Not supported although a hierarchical version of PIM has been discussed

5.3 Sparse Group Distribution: PIM-SM vs CBT

One major difference between PIM-SM and CBT is that the former has the capability to change from a shared tree to a shortest-path tree. Comparisons to CBT for both types of PIM-SM trees, i.e., shared and shortest-path, are presented in this section.

The results of a simulation done by R. Voigt (Naval Postgraduate School) and presented at a meeting of the IETF IDMR working group in 1995 and also summarized in [1] are as follows:

- PIM-SM (shortest-path trees) improves the delay of CBTs by 5 to 20%, but incurs about double the overhead of CBT; and,
- PIM-SM (shared tree) consistently shows much longer delays (>50%) than CBT and incurs about the same overhead as CBT.

Ballardie in [1] concludes from these results and other simulation results obtained by D. Estrin and L. Wei in [35] that PIM appears to be a superset of CBT and that there are distinct advantages to using CBT because:

- the end-to-end delay is only slightly lower (10% on average) with PIM-SM (shortest-path trees) than the CBT;

- in shared-tree mode, there is no advantage to using PIM-SM over CBT;
- in terms of bandwidth utilization/overhead, both protocols are shown to be about equal; and,
- CBTs offer the most favourable scaling characteristics for the typical case where there are some senders within or outside a group of receivers.

A more recent performance and resource cost comparison study for the CBT and PIM protocols is that of Tom Billhartz (Harris Corporation in Melbourne, Florida) and his colleagues. Billhartz obtained the results [36] summarized in Table 5 using the OPNET network simulation tool.

TABLE 5. Comparison of CBT and PIM Protocols

Metric	**Protocol**			
	CBT	**PIM-SM (shared tree)**	**PIM-SM (source-based)**	**PIM-DM**
End-to-End Delay	Low	Low	Low	Low
Network Resource Usage (number of hops travelled by all copies of data packets)	Moderate	Moderate	Moderate	High in sparse group environment
Overhead Traffic Percentage	Small in the cases simulated but proportional to number of joins per second	Small in the cases simulated but proportional to number of joins per second	Small in the cases simulated but proportional to number of joins per second	High
Join Time	Low	Low	Low	High on average
Traffic Concentration	High	Highest	Low	Lowest
Routing Table Size	Linear with the number of groups	Proportional to the product of number of groups and mean number of senders per group	Proportional to the product of number of groups and mean number of senders per group	Proportional to the product of number of groups and mean number of senders per group
Implementation Difficulty	Low to Moderate	Complex	Complex	Moderate

Billhartz's conclusions are similar to those of Voigt, Estrin/Wei and Ballardie mentioned above:

- PIM-SM source-based tree (and PIM-DM) has slightly lower delays than CBT and PIM-SM shared-tree, but the absolute delays are all very small;
- PIM-SM source-based and PIM-DM deliver packets 12 to 31% faster than CBT, depending on the topology simulated;
- Network resource usage is similar for all of the protocols except PIM-DM, which periodically floods data on the network;
- Traffic concentration is observed in CBT and PIM-SM (shared-tree), but does not degrade performance significantly; and,
- CBT has the lowest overhead percentage of the protocols examined, approximately 0.3%.

According to Billhartz, the size of the routing table and the impact of the timers on the operating system may become a major factor. Operation of PIM-SM (shared-tree) or PIM-DM for a large number of members and groups requires each router to maintain large routing tables. On average, PIM-SM (shared-tree) routers have fewer routing table entries and fewer timers than the source-based tree protocols. CBT routers have even fewer. Based on these observations, Billhartz argues that with current technology, "... CBT is the best suited multicast protocol for environments with a large number of groups, each with many senders."

Another major difference between PIM-SM and CBT is that PIM-SM uses "soft-state" whereas CBT uses "hard-state". A negative consequence of the "hard-state" approach is that CBT branches do not automatically adapt to underlying multicast route changes. This is in contrast to the "soft-state" or data-driven approach — data always follows the path as specified in the routing table. Provided reachability is not lost, it is advantageous, from the perspective of uninterrupted packet flow, that a multicast route be kept constant, but the two disadvantages are that a route may not be optimal for its entire duration, and "hard-state" requires the incorporation of control messages that monitor reachability between adjacent routers on the multicast tree. Unless some form of message aggregation is employed, this control message overhead can be quite considerable, especially when changes need to be detected quickly.

5.4 Exterior Protocol: BGMP vs Wide-Area Multicast Protocols

The designers of BGMP are proposing a global, Internet-wide multicast routing infrastructure. In this architecture, BGMP is run by domain border routers to construct an inter-domain bidirectional shared tree for a group while allowing any existing multicast routing protocol such as DVMRP, PIM and CBT to be used within individual domains. Such intra-domain multicast routing protocols are also known as Multicast Interior Gateways Protocols (MIGPs).

BGMP is the only "Multicast Exterior Gateways Protocol" currently being considered by the IDMR working group. Potential contenders would most likely be the hierarchical versions of the existing MIGPs, i.e., HPIM, OCBT/HIP, and Hierarchical DVMRP.

The non-hierarchical MIGPs do not compare well with BGMP. The dense group distribution protocols lack scalability. They do not scale well to groups that span the Internet because of periodic flooding of data packets (DVMRP and PIM-DM) or group membership information to all the routers (MOSPF) throughout the network. The sparse group distribution protocols like CBT and PIM-SM scale better but lack hierarchism and autonomy. For instance, the mechanism for distributing the mapping of a group to its corresponding core (Rendezvous Point) router requires flooding of the set of all routers that are willing to be cores. Hierarchism is needed to provide a means of routing between heterogeneous domains that might use any multicast protocol internally. Also, since inter-domain routing involves the use of resources in autonomously administered domains, the policy constraints of such domains needs to be accommodated.

As described in Section 3.1 while discussing desirable properties of multicast routing protocols, there are open issues in the design of an Internet-wide multicast routing architecture that require further research.

6.0 Other Issues and Properties

6.1 IGMP

The Internet Group Management Protocol (IGMP) is the protocol used by IP systems to report their IP multicast group memberships to neighbouring multicast routers. Strictly speaking, IGMP is not a routing protocol but an integral part of the standard IP network-level protocol. It is mentioned here because IGMP is occasionally enhanced by designers to carry information specific to the multicast routing protocols being proposed.

IGMP is implemented directly over IP and has only two messages: "host membership query" and "host membership reports." Both messages have the same format with the type field being set to 1 in membership queries and to 2 in reports. Other type values are used to implement DVMRP or PIM which do not have their own protocol headers.

There are three versions of IGMP:

- IGMPv1, defined in RFC 1112 [37], is the first widely-deployed version and the first version to become an Internet official protocol standard (STD 5);
- IGMPv2, specified in RFC 2236 [38], is currently a Proposed Standard and will add "low leave latency," i.e., will allow group membership termination to be quickly reported to the routing protocol, which is important for high-bandwidth multicast groups and/or subnets with highly volatile group membership; and,
- IGMPv3, an Internet-Draft [39] at the moment, adds support for "source filtering", that is, the ability for a system to report interest in receiving only from specific source addresses, or from all but specific source addresses, sent to a particular multicast address.

Newer versions are designed to be interoperable with older versions.

In addition, the IDMR working group of the IETF is looking at two other possible upgrades of IGMP:

1. **Domain Wide Multicast Group Membership Reports (DWRs)** — DWR is a group membership protocol at the domain level. When using a hierarchical multicast routing protocol like Hierarchical DVMRP or BGMP, the inter-domain protocol needs to learn of group memberships inside domains. Although some intra-domain routing protocols can provide this information easily to the domain border routers, some cannot. In DWR, packets are sent as IGMP packets to allow group membership inside a domain to inform in a protocol-independent fashion. DWR specifies a behaviour that can be used with any intra-domain protocol, along with optimization for certain intra-domain protocols (MOSPF, PIM, CBT), and a transition scheme so that all interior routers need not be updated. DWR is currently an Internet-Draft [40].
2. **IGMP Multicast Router Discovery** — A method for discovering multicast capable routers is necessary for layer-2 bridging devices. Currently, IGMP "group membership query" message is inadequate for discovering multicast routers because query messages are suppressed once one querier gets elected. In order to "discover" multicast routers, two

new types of IGMP messages are proposed: Multicast Router Advertisement and Multicast Router Solicitation. These two messages can be used by any layer-2 device that listens to IGMP to find multicast routers to determine where to send multicast source data and IGMP host membership reports. This proposal is currently an Internet-Draft [41].

6.2 Interoperability

Currently, there is no inter-domain multicast routing protocol standard approved by the IETF. For the time being, BGMP is an Internet-Draft and this puts certain topological constraints on the multicast infrastructure development. In the interim, the IDMR working group has issued an Internet-Draft [42] describing rules to allow efficient interoperation among multiple independent multicast routing domains. Specific instantiations of these rules are given for the DVMRP, MOSPF, PIM-DM, and PIM-SM multicast routing protocols, as well as for IGMP-only links.

As the Internet multicast infrastructure evolved from the initial MBONE architecture, many routers today support DVMRP. This predominant protocol is treated by many as the *de facto* standard for constructing delivery trees for inter-region multicast routing. For this reason, protocol specification for wide-are multicast routing (e.g. PIM-SM and CBT) ensure direct interoperability with DVMRP.

Interoperating with DVMRP means that simple policy routing can be achieved by manipulating the DVMRP metrics at a region's boundary. Note however that DVMRP requires these metrics to be symmetric between the border router pairings. If different inter-region multicast protocols were used, metric translation between protocols would be required that would make the interoperability situation more complex.

6.3 Asymmetry and Directivity

Most Internet unicast and multicast routing protocols have been designed for optimum performance assuming bidirectional symmetric communication links. Symmetry means that links have the same characteristics in each direction. There are specific problems in the case of asymmetric (unidirectional, if the bandwidth is zero in one direction) links. In general, three cases for unidirectional and/or asymmetric links may be envisaged:

1. unidirectional links on top of bidirectional underlying network (wired Internet);
2. bidirectional islands connected via unidirectional links; and,
3. the general case of asymmetric and possibly unidirectional links.

The UniDirectional Link Routing (UDLR) working group of the IETF was set up to provide a solution to the first case listed above. An example of such a configuration is a broadcast satellite network where the receivers are equipped with receive only antennas and terrestrial back channels, i.e., cables, and low speed modems for the return traffic to the feed[1]. Another common example might be wireless networks where pairs of routers do not have the same transmitting power. To solve the problem, the group is experimenting with two different approaches. Both approaches support IP multicast datagrams forwarding over the unidirectional link.

In the first approach, the group produced Internet-Drafts for RIP, OSPF and DVMRP where modifications are defined to make routing and multicasting communications over asymmetric links feasible. In the case of DVMRP, the group realized that the current implementation of DVMRP, *mrouted*, does not accurately follow the RFC; the group had to rely on the C code itself to see what modifications could be made. Needless to say, changes being proposed in [43] by the group to support unidirectional links may need to be reviewed once version 3 of DVMRP becomes available.

In the second approach, tunnels are dynamically set up as a virtual back-channel of the unidirectional link to carry routing information between the broadcast feeders and the wired network. This solution has the advantage that no modification of the current routing protocols or their implementations is necessary. The tunnelling approach adds a layer between the network interface and the routing software on both ends of the unidirectional link (or between some intermediate routers), resulting in the emulation of a bidirectional link where only a unidirectional link is available. The approach is described in [44] and specified in details in an Internet-Draft. The latter document [45] includes a *Dynamic Tunnel Configuration Protocol (DTCP)* which is designed to provide a means for receivers to dynamically discover the presence of feeds and to maintain a list of operational tunnel end-points. It is based on feed periodical announcements over the unidirectional link which contain tunnel end-point addresses. Receivers listen to these announcements and maintain a list of tunnel end-points.

The tunnelling solution is preferred to the routing protocol modification approach because it can quickly be implemented by satellite operators and service providers whom are anxious to offer Internet access via satellite. Unfortunately, this transparent short term solution does not imply optimal operation.

Finally, the UDLR group is also developing a new routing protocol [46] called *Circuit Discovery Unidirectional Link Routing Protocol* to provide a long term solution to the problem of dynamic routing in a unidirectional network.

All of the work described in this section is experimental and requires further research. One should stress the urgency of this work since integration of satellite networks (LEO, MEO, GEO) with the Internet is likely to play an increasingly important role in the not so distant future as evidenced by the large number of systems planned or in operation, such as GPS, ACTS, DirecTV, Iridium, DirecPC, SpaceWay, and Teledesic. Internet Service Providers (ISPs) relying on cable and satellite communications infrastructures in conjunction with increasing amounts of audio, video, and other content types are starting to build new business models and introduce value-added services that include IP multicast.

[1] A router connected to an unidirectional link with a send-only interface.

As more content is carried by these ISPs, they are beginning to realize how crucial IP multicast is to them. The use of cable and satellite communications infrastructures is one of the key areas to watch right now. Several of the ISPs are already coming up with their own sub-optimal unicast-based distribution methods.

6.4 Security

An Internet Security Architecture [47] has been proposed by the IPSEC working group of the IETF. The proposed security mechanisms are implemented at the network layer. Providing security services for multicast, such as traffic integrity, authentication, and confidentiality, requires securely distributing a group (session) key to each of a group's receivers. However, multicast key distribution is not addressed to any significant degree by the Internet Security Architecture [47].

Recently, the Group Key Management Protocol (GKMP) [48], was proposed as an experimental protocol for the management of cryptographic keys for multicast communications. The protocol was initially designed for military networks (command and control and weapons control systems), but has been adapted to the Internet [49]. GKMP does not rely on a centralised key distribution centre, a method which would not scale for wide-area multicasting, but rather places the burden of key management on a group member(s). In short, GKMP creates keys for cryptographic groups, distributes keys to group members, ensures (via peer to peer reviews) rule based access control of keys, denies access to known compromised hosts, and allows hierarchical control of group actions. GKMP is an application layer protocol that is independent of the underlying communication protocol. However, if multicast service is available, it will speed the rekey of the cryptographic groups. Hence, GKMP does use multicast services if they are available.

Many network layer multicast protocols, such as DVMRP, MOSPF and PIM for instance, do not have their own protocol header(s), and so cannot provide their own security mechanism; they must rely on whatever security is provided by IP itself. One exception is CBT. RFC 1949 [50] defines an experimental protocol called "*Scalable Multicast Key Distribution*" where CBT is used (solely) for multicast key distribution. It is said that CBT can optionally provide for the secure joining to a CBT group tree. Furthermore, it does not preclude the use of other multicast protocols for the actual multicast communication itself; that is, CBT need only be the vehicle to distribute the keys. As for GKMP, the scheme does not require a centralised key distribution centre. Instead, each group has its own group key distribution centre (GKDC) and the functions it provides are "passed on" to other nodes as they join the tree. A CBT primary core initially take on the role of a GKDC.

The work on cryptographic keys for multicast communications is experimental and requires further research.

6.5 Multicast Scoping

When multicast applications operate in the Internet, it is clear that not all groups should have global scope. Being able to constrain the scope of a session allows the same multicast address to be in use at more than one place provided the scope of the sessions does not overlap. Most current IP multicast implementations confine the distribution of multicast traffic, i.e., achieve "scoping," by using the Time-To-Live (TTL) field in the IP header. In addition, the TTL is also used in its traditional role to limit datagram lifetime. Given these often conflicting roles, TTL scoping has often been complex and difficult to implement. There are other architectural problems with TTL scoping. One concerns the interaction of TTL scoping with broadcast and prune protocols like DVMRP and PIM-DM. In many cases, TTL scoping can prevent pruning from being effective: the router which discards the packet will not be capable of pruning any upstream sources. Another problem is that there are circumstances where it is difficult to consistently choose TTL thresholds to achieve the desired scoping. For instance, it is impossible to configure overlapping scope regions as shown in Figure 1.

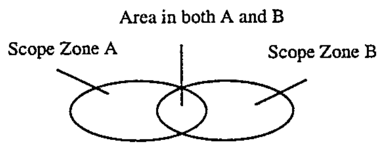

FIGURE 1. Overlapping scope zones possible with administrative scoping

An alternative to TTL scoping is administrative scoping. Administrative scoping allows the configuration of a boundary by specifying a range of multicast addresses that will not be forwarded across that boundary in either direction. To avoid introducing significant address management complexity, the multicast addressees may be dynamically allocated.

The Multicast-Address Allocation (MALLOC) working group of the IETF has recently proposed a hierarchical dynamic multicast address allocation architecture for the Internet [51]. This architecture assumes that the primary scoping mechanism in use is administrative scoping and that TTL scoping will cease to be used before the architecture is used widely. There are three parts to this architecture:

1. A *Multicast address allocation based on the Dynamic Host Configuration Protocol (*MDHCP) [52] that a multicast client uses to request a multicast address from a local multicast address allocation server (MAAS);
2. A multicast *Address Allocation Protocol (*AAP) [53] that MAAS servers use to claim multicast addresses and inform their peer MAAS servers which addresses are in use; and,
3. A *Multicast Address-Set Claim (MASC)* protocol [30] that allocates multicast address sets to domains using a "listen and claim" with collision detection approach. Individual addresses are allocated out of these sets by MAAS servers.

MASC is performed by routers that run BGP4+. The BGP4+ protocol serves as the glue between MASC and BGMP (or other inter-domain multicast tree construction protocol). The address sets can be used by the latter to construct inter-domain group-shared trees. This implies that MASC itself cannot use multicast and must rely on unicast TCP, in parallel to those TCP connections used by BGP4+.

MASC domains form a hierarchy that reflects the structure of the inter-domain topology. It is expected that allocation domains will normally coincide with unicast autonomous systems.

6.6 Quality of Service

PIM and CBT would need extensive modifications to accommodate the QoS requirements of the emerging multimedia services. MOSPF is more suited for QoS-based multicast routing since it can gather the network-wide QoS information from the underlying link-state routing algorithm. A series of extensions to OSPF and MOSPF have recently been proposed in an Internet-Draft [54] to provide QoS routing in conjunction with a resource reservation protocol such as RSVP. The extensions, called *Quality of Service Path First Routing* (QOSPF), generate advertisements indicating the resources available and the resources used. They are sent within the OSPF routing domain and the paths are computed based on topology information, link resource information, and the resource requirements of a particular data flow.

Protocols like QOSPF do not address several other important QoS issues tied to the Internet infrastructure and can only be part of a broader QoS framework. To date, most of the work on QoS has been within the context of individual architectural layers such as the distributed system platform, operating system, layer-4 transport subsystem and layer-3 network. It can be argued that an end-to-end approach should be adopted to meet application level QoS requirements. This view is shared by the authors in [55] where the state-of-the-art in the development of QoS architectures has recently been examined in detail. The IETF has also recognized that in order to support the new multimedia applications, the Internet's "best-effort" service model needs to be enhanced. The Integrated Services (intserv) working group has been set up to define new service classes such as Guaranteed Service Class and Controlled Load Service Class. The goal is to specify a minimal set of global requirements to help transition the Internet into a robust integrated-service communications infrastructure.

The subject is vast and beyond the scope of this report. It is, however, suffice to say that in a packet-switched environment like the Internet, there are at least five major architectural components that try, individually or in combination, to meet the QoS requirements in an efficient manner: Flow Specification, Routing, Resource Reservation, Admission Control and Packet Scheduling. In a recent PhD thesis [56] by S. Verma at the University of Toronto, QoS routing and resource reservation are shown to be interdependent and sha-reing the same goal of achieving efficient utilization of resources. This interaction seems to have been ignored in the existing literature. Verma's dissertation shows that these two aspects must be addressed at the same time to achieve the best performance.

Given the complexity of the issues, the bulk of the efforts spent today by most multicast protocol designers is in preparation for the new network architectures, such as IPv6, where the integration of QoS components and multicasting should be somewhat easier to develop.

6.7 Tactical Environment

When considering multicast routing architectures for use in a military environment most, if not all, of the trade-offs mentioned so far apply. Among the issues being investigated by various groups of the IETF, two should be noted that are particularly relevant to the tactical communication network environment:

1. The need for security mechanisms to provide routing/data traffic integrity, authentication and confidentiality as well as to control access to group membership (Section 6.4);
2. The need for unidirectional link routing protocols (Section 6.3) when operating in harsh conditions where tactical links may be non-symmetric, simplex-only and/or of low bandwidth.

In addition, multicast routing protocols for the tactical environment require three other important characteristics; mainly, robustness, adaptability and reliability, which are now briefly summarized.

3. Robustness — It has been said that the shared tree approach provides superior scalability than the source based tree approach. However, the shared delivery tree may suffer from the traffic concentration problem. That is, the core (or rendezvous point) of a particular group can potentially become a traffic "hot-spot" or "bottleneck". Worst, it can momentarily become a "traffic-trap" for the whole group in case of server failure[1]. The source-based tree is more robust (less vulnerable) because "hot-spots" are less likely to occur due to the distributed nature of the routing tree. Because of that same characteristic, the source-based tree can also provide better immunity to non-reliable links where routing information may be lost due to a noisy propagation channel, radio interference or jamming affecting a bounded area of the network.
4. Adaptability — Most current multicast routing protocols are able to adapt to change in the network topology. This is a desirable property because connectivity between routers can be maintained when links change state, for instance, when links are added, removed or reconfigured. Unfortunately, none of the multicast protocols today are able to adapt quickly enough to such changes. Whenever a change occur, routers enter a transient period which limits how often re-configurations can take place. In a military deployment where mobility of hosts/routers is a prime requirement, frequent re-configuration of network, specially on short notice, forces the network to react and adjust quickly. The per-

[1] This transient condition can be minimized by configuring a small strategic ordered set of alternative RPs.

formance of present day multicast routing protocols are less than satisfactory because the network cannot keep up with the required rate of change: a time lag can be observed in the exchange of the multicast routing information throughout the network.

5. Reliability — Since IP is a best-effort delivery protocol, reliable data transfer to specific multicast groups cannot be achieved unless explicit acknowledgement/retransmission procedures are implemented. As mentioned in the introduction of this paper, this part is best handled by designing reliable multicast protocols operating above the network layer since data reliability involves end-to-end data packet delivery. As for the reliability of the multicast routing information itself, on a hop-by-hop basis, two basic approaches are commonly used:

   i) information is repeated at regular intervals without being acknowledged by the destination, e.g., PIM-SM; and,

   ii) information is explicitly acknowledged and only repeated if a timer expires before any acknowledgement has been received, e.g., CBT.

   The second approach requires less message overhead but does not adapt automatically to underlying multicast route changes as reachability messages are needed between adjacent routers on the multicast tree to rapidly detect changes.

7.0 Summary

The state of the multicast routing technology is perhaps best summarized by Maufer [59]:

> "The research community has thus far created a number of multicast routing protocols, each of which may be applicable in different scenarios; however, no 'one-size-fits-all' multicast routing protocol has yet been invented".

In fact, the trade-offs between the various protocols available today are, for the most part, closely related to the position each protocol occupies in the multicast routing protocol family tree as shown below.

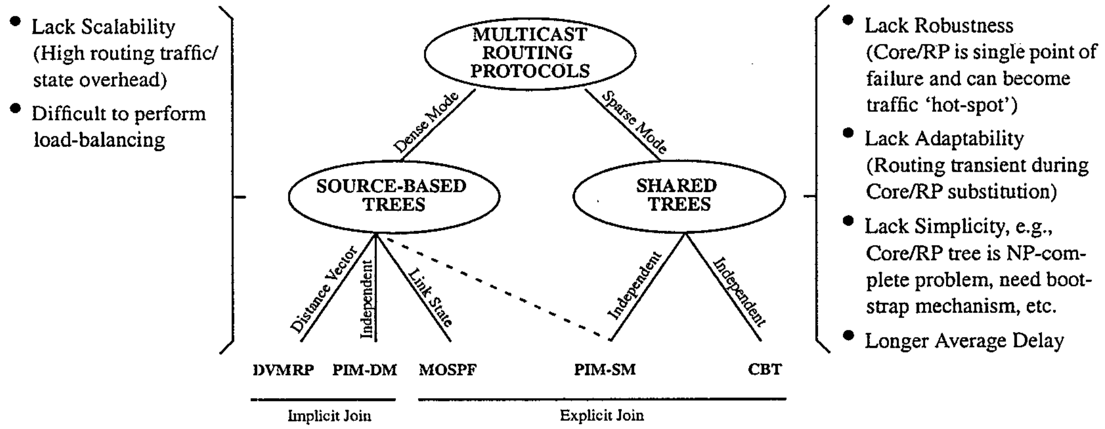
• Lack Scalability (High routing traffic/ state overhead)
• Difficult to perform load-balancing
MULTICAST ROUTING PROTOCOLS
Dense Mode
Sparse Mode
SOURCE-BASED TREES
SHARED TREES
Distance Vector
Independent
Link State
Independent
Independent
DVMRP
PIM-DM
MOSPF
PIM-SM
CBT
Implicit Join
Explicit Join
• Lack Robustness (Core/RP is single point of failure and can become traffic 'hot-spot')
• Lack Adaptability (Routing transient during Core/RP substitution)
• Lack Simplicity, e.g., Core/RP tree is NP-complete problem, need bootstrap mechanism, etc.
• Longer Average Delay

FIGURE 2. Multicast Routing Protocol Family Tree

Sparse Mode protocols trade off using bandwidth liberally, which is valid in a densely populated intranet/LAN environment where Dense Mode protocols operate best, for techniques that are much better suited for large WANs, where bandwidth is scarce and expensive.

One should note that none of the existing multicast routing protocols is likely to perform satisfactorily in military environment. This is due to the fact that none of them was specifically designed for the wireless and low bandwidth environment that is prevalent in military networks, where link symmetry cannot always be achieved despite being needed for proper operation. Furthermore, most of the protocols require better robustness, adaptability and reliability characteristics.

A summary of the trade-offs, benefits and deficiencies of the five most popular multicast routing protocols described in this document are given in Table 6 and Table 7.

TABLE 6. Summary of DVMRP, MOSPF and PIM-DM

Metric	Protocol		
	DVMRP	MOSPF	PIM-DM
Protocol Status	Experimental Internet Protocol	Proposed Internet Standard	Internet-Draft
Popularity	High. Selected by most to interoperate with sparse group distribution protocols like CBT and PIM-SM.	Moderate	Emerging. Used to be a proprietary protocol defined by Cisco Systems Inc. Currently supported by few router vendors.
Implementation Difficulty	Very simple protocol	Complex protocol	Very simple protocol
Interoperability	High. Capable of interconnecting domains running other multicast routing protocols (Interoperability supported by virtually all vendors of multicast-capable routers).	Low. Eliminate all non-MOSPF routers when building delivery tree. Cannot run in non-OSPF domains (if intranet is not 100% OSPF, PIM-DM may be a better choice). Relies on DVMRP to interconnect with domains running other multicast routing protocols.	Low. Relies on DVMRP to interconnect with domains running other multicast routing protocols.
Routing Technology	Distance Vector. Reacts slowly to changes and prone to routing loops. Limited network diameter: 15 hops. Must maintain source-specific state when not on-tree: 1) Keeping state in off-tree routers is a waste of valuable router memory; and, 2) Source-specific state doesn't scale well as the number of sources increases. Multicast traffic is periodically broadcast across the entire internetwork.	Link State. Reacts quickly to changes, loop-free, but computationally intensive. Calculation of source-based SPF tree in memory may cause possible strain on CPU resources when many new groups appear at about the same time. Does not support tunnels. Relatively easy to overwhelm the routers in an area by maliciously spraying in multicast packets with randomly chosen (source,group) pairs.	No built-in unicast routing protocol. Relies on underlying unicast routing protocol capability. Incurs more overhead than DVMRP, including some possible excess packet duplication.
Handling of non-multicast capable routers	User manually sets up tunnels to reach multicast capable routers	Does not support tunnels. Automatically eliminates all non-multicast OSPF routers from the SPF tree. Unicast and multicast connectivity to a destination may not follow the same path	Relies on underlying unicast routing protocol capability
Underlying unicast routing requirement	None	Multicast extensions built on top of OSPF	Any unicast routing protocol
QoS support	1 metric through tunnelling	5 types of service	None. Assumes that point-to-point routes are symmetric
Hierarchism	Not supported although a hierarchical version of DVMRP has been proposed.	2 Levels	Not supported although a hierarchical version of PIM has been proposed

TABLE 7. Summary of CBT and PIM-SM

Metric	Protocol	
	CBT	PIM-SM
Protocol Status	Experimental Internet Protocol	Experimental Internet Protocol
Implementation Complexity Level	Low to Moderate	High
Interoperability	High. Protocol Independent.	High. Protocol Independent.
Routing Technology	No built-in unicast routing protocol. Relies on underlying unicast routing protocol capability. Bidirectional state, data may flow in either direction along a branch (unique feature to CBT).	No built-in unicast routing protocol. Relies on underlying unicast routing protocol capability. Unidirectional state, data may only flow away from the RP, not toward it. Can use either source-based or shared trees.
Routing Table Size	Linear with the number of groups. Amount of state information is invariant with the number of sources	Proportional to the product of number of groups and the mean number of senders per group
Traffic Concentration	High	Highest (shared tree); Low (source-based).
End-to-End Delay	Low	Low (shared tree); Lowest (source-based).

8.0 Conclusion

The benefits of multicasting are becoming evermore apparent, and its use more widespread. IP multicast enables many new types of applications and reduces network loads.

Since 1996, the IP Multicast Initiative[1] (IPMI) has been promoting the deployment of industry-standard IP multicast technology. Launching IP multicasting into the mainstream is however taking longer than expected. Despite broad industry backing and support by many vendors of network infrastructure elements such as routers, switches, network interface cards and application software, definition of the IP multicast architecture lags behind the technology and causes transitional approaches to be used. Advances are being made in several areas but much experimental work remains to be done before an Internet-wide deployment becomes truly functional. Some of the open issues include:

1. **Policy/QoS Support** — Multicast policy and access-control are nearly nonexistent. Route filtering and packet filtering are the principle means today to achieve crude forms of policy routing. Whereas existing Internet multicast routing protocols do not consider QoS metrics, multimedia applications are usually sensitive to delay, reliability, and bandwidth availability. Furthermore, it is unclear how much longer the free funding model (until recently, most networks were funded as "public goods") of the Internet will last. In such a context, support for Policy/QoS could become an important requirement. Despite the Internet community's reluctance to invest in comprehensive Policy/QoS routing, due to its complexity, it can be assumed that this issue, pioneered by such concepts as Inter-Domain Policy Routing (IDPR) [10] and Source Demand Routing Protocol (SDRP) [11], will have to be re-examined. At this time, there are too many unresolved items in the multicast infrastructure to see Policy/QoS get an appropriate amount of attention. This development may have to wait for improvements to IP itself as promised by IPv6.
2. **Autonomy**— A fully IP multicast-enabled Internet requires an IDMR standard protocol. PIM, CBT, MOSPF and DVMRP are not designed for multiple autonomous systems and cannot limit the propagation of routing information based on policies and rules that administrators might want to use. The IETF has yet to develop a single Internet standard for inter-domain multicast routing; however, BGMP is the prime contender.
3. **Hierarchism** — The growth of IP multicast is severely limited if all routers must maintain all the routing information for the whole network. The only way to hide information is with a hierarchical routing topology[2]. It is interesting to note that the same solution is likely to make multicast address allocation scale and solve the multicast scoping problem. The IETF has yet to converge on an Internet standard for intra-domain multicast routing. In fact, two standards, one for dense mode (e.g. MOSPF) and one for sparse mode (e.g. OCBT or another hierarchical version of CBT), may be necessary to accommodate the full range of multicast applications.

[1] A multi-vendor cooperative group managed by Stardust Technologies Inc.

[2] Kamoun and Kleinrock have shown in [57] that the optimal number of levels for an *N* router network is *ln(N)*, requiring *eln(N)* entries per router.

4. **Directivity** — Integration of satellite networks with the Internet is on its way. ISPs relying on cable and satellite communications infrastructures are starting to build new business models and introduce value-added services that include IP multicast. Existing Internet unicast and multicast routing protocols have been designed for optimum performance assuming bidirectional symmetric communication links. Proposed solutions such as back channel through tunelling are sub-optimal and should be used in the short-term only.

As described, deploying multicast routing algorithms in a very large network is a complex task. Further research is required before a comprehensive multicast routing protocol architecture can be defined that meets the requirements in either a civilian or military network.

9.0 References

[1] Ballardie, T. and Crowcroft, J. (1995). *Core Based Tree (CBT) Multicast — An Analysis of Multicast Routing Architectures*, IEEE/ACM Transactions on Networking.

[2] Hedrick, C. (1988). *Routing Information Protocol*, RFC 1058, Rutgers University.

[3] Moy, J. (1998). *OSPF Version 2*, RFC 2328.

[4] Hedricks, C.L. (1991). *An introduction to IGRP,* Center for Computer and Information Services, Laboratory for Computer Science Research, Rutgers University.

[5] Farinacci, D. (1993). *Introduction to Enhanced IGRP (EIGRP)*, Cisco Systems Inc.

[6] Oran, D. (1990). *OSI IS-IS Intra-domain Routing Protocol*, RFC 1142.

[7] ISO/IEC JTC 1 (1990). *Intermediate System to Intermediate System Intra-Domain Routeing Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)*, ISO-DP-10589.

[8] Rekhter, Y. and Li, T. (1995). *A Border Gateway Protocol 4 (BGP-4)*, RFC 1771.

[9] ISO/IEC (1993). *Information Processing Systems - Telecommunications and Information Exchange between Systems - Protocol for Exchange of Inter-domain Routeing Information among Intermediate Systems to Support Forwarding of ISO 8473 PDUs*", ISO/IEC IS10747

[10] Streenstrup, M. (1993). *Inter-Domain Policy Routing Protocol Specification: Version 1*, RFC 1479.

[11] Estrin, D., et al. (1996). *Source Demand Routing: Packet Format and Forwarding Specification (Version 1)*, RFC 1940.

[12] Perkins, C. (1996). *IP Encapsulation within IP,* RFC 2003.

[13] Postel, J. and Reynolds, J. (1998). *Internet Official Protocol Standards*, RFC 2400.

[14] Wall, D. (1980). *Mechanisms for Broadcast and Selective Broadcast*, PhD thesis, Stanford University, California, U.S.A.

[15] Waitzman, D., Partridge, C., and Deering, S. (1988). *Distance Vector Multicast Routing Protocol*, RFC 1075.

[16] Moy, J. (1994). *Multicast Extensions to OSPF,* RFC 1584.

[17] Deering, S. et al. (1998). *Protocol Independent Multicast Version 2 Dense Mode Specification*, Internet-Draft.

[18] Estrin, D. et al. (1998). *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*, RFC 2362.

[19] Ballardie, A. (1997). *Core Based Trees (CBT version 2) Multicast Routing*, RFC 2189.

[20] Thaler, D., Estrin, D., and Meyer, D. (1998). *Border Gateway Multicast Protocol (BGMP): Protocol Specification*, Internet-Draft of the IETF IDMR working group.

[21] Coltun, R. et al. (1998). *DVMRPv1 Applicability Statement for Historic Status*, Internet-Draft of the IETF IDMR working group.

[22] Pusateri, T. (1998). *Distance Vector Multicast Routing Protocol*, Internet-Draft of the IETF IDMR working group.

[23] Thyagarajan, A.S., and Deering, S.E. (1995). *Hierarchical Distance-Vector Multicast Routing for the MBone*, Proc. of ACM SIGCOMM'95, 25(4), 60-66

[24] Estrin, D. et al. (1997). *A Dynamic Bootstrap Mechanism for Rendezvous-based Multicast Routing*, Technical Report USC CS TR97-644, University of Southern California.

[25] Handley, M. and Crowcroft, J. (1995). *Hierarchical Protocol Independent Multicast (HPIM).*

[26] Ballardie, A. (1998). *Core Based Trees (CBT Version 3) Multicast Routing — Protocol Specification*, Internet-Draft of the IETF IDMR working group.

[27] Chang, Y.C. (1996). *Core Group Based Tree: A Hierarchical Multicast Routing Protocol*, Minutes of the 35th IDMR IETF meeting.

[28] Shields, C. and Garcia-Luna-Aceves, J.J. (1997). *The Ordered Core Based Tree Protocol*, Proc. IEEE INFOCOM 1997, Kobe, Japan.

[29] Shields, C. and Garcia-Luna-Aceves, J.J. (1998). *The HIP Protocol for Hierarchical Multicast Routing*, Proc. 17th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC'98), Puerto Vallarta, Mexico.

[30] Estrin, D. et al. (1998). *The Multicast Address-Set Claim (MASC) Protocol*, Internet-Draft of the IETF MALLOC working group.

[31] Banerjea, A., Faloutsos, M., and Pankaj, R. (1998). *Designing QoSMIC: A Quality of Service sensitive Multicast Internet protoCol*, Internet-Draft of the IETF IDMR working group.

[32] Hodel, H. (1998). *Policy Tree Multicast Routing: An Extension to Sparse Mode Source Tree Delivery*, ACM SIGCOMM, 28(2).

[33] Parsa, M. and Garcia-Luna-Aceves, J.J. (1997). *A Protocol for Scalable Loop-Free Multicast Routing*, IEEE Journal on Selected Areas in Communications, 15(3), 316-331

[34] Lui, C., Lee, M. J., and Saadawi, T. N. (1997). *A Scalable Multicast Routing*, Proc. MILCOM'97.

[35] Estrin, D. and Wei, L. (1994). *The Trade-Offs of Multicast Trees and Algorithms*, In International Conference on Computer Communications and Networks.

[36] Billhartz, T. et al. (1997). *Performance and Resource Cost Comparisons for the CBT and PIM Multicast Routing Protocols*, IEEE Journal on Selected Areas in Communications, 15(3), 304-315

[37] Deering, S. (1989). *Host Extensions for IP Multicasting*, RFC 1112 (Internet Standard).

[38] Fenner, W. (1997). *Internet Group Management Protocol, Version 2*, RFC 2236 (Proposed Standard).

[39] Cain, B., Deering, S., and Thyagarajan, A. (1997). *Internet Group Management Protocol, Version 3*, Internet-Draft of the IETF IDMR working group.

[40] Fenner, W. (1998). *Domain Wide Multicast Group Membership Reports*, Internet-Draft of the IETF IDMR working group.

[41] Biswas, S. and Cain, B. (1998). *IGMP Multicast Router Discovery*, Internet-Draft of the IETF IDMR working group.

[42] Thaler, D. (1998). *Interoperability Rules for Multicast Routing Protocols*, Internet-Draft of the IETF IDMR working group.

[43] Duros, E. and Dabbous, W. (1997). *Handling of unidirectional links with DVMRP*, Internet-Draft of the IETF UDLR working group.

[44] Dabbous, W., Duros, E., and Ernst, T. (1997). *Dynamic Routing in Networks with Unidirectional Links", Proc. of the 2nd International Workshop on Satellite-based Information Services*, Budapest, Hungary.

[45] Duros, E. et al. (1998). *A Link Layer Tunneling Mechanism for Unidirectional Links*, Internet-Draft of the IETF UDLR working group.

[46] Ernst, T. and Dabbous, W. (1997). *A Circuit-based Approach for Routing in Unidirectional Link Networks*, INRIA Sophia-Antipolis, France. Submitted to INFOCOM'98

[47] Atkinson, T. (1995). *Security Architecture for the Internet Protocol*, RFC 1825.

[48] Harney, H. and Muckenhirn, C. (1997). *Group Key Management Protocol (GKMP) Specification*, RFC 2093 (Experimental Protocol).

[49] Harney, H. and Muckenhirn, C. (1997). *Group Key Management Protocol (GKMP) Architecture*, RFC 2094 (Experimental Protocol).

[50] Ballardie, A. (1996). *Scalable Multicast Key Distribution*, RFC 1949 (Experimental Protocol).

[51] Handley, M., Thaler, D., and Estrin, D. (1997). *The Internet Multicast Address Allocation Architecture*, Internet-Draft of the IETF MALLOC working group.

[52] Patel, B.V., Shah, M., and Hanna, S.R. (1998). *Multicast address allocation based on the Dynamic Host Configuration Protocol (MDHCP)*, Internet-Draft of the IETF MALLOC working group.

[53] Handley, M. (1998). *Multicast Address Allocation Protocol (AAP)*, Internet-Draft of the IETF MALLOC working group.

[54] Zhang, Z. et al. (1997). *Quality of Service Extensions to OSPF or Quality of Service Path First Routing (QOSPF)*, Internet-Draft.

[55] Aurrecoechea, C., Campbell, A. T., and Hauw, L. (1998). *A Survey of QoS Architectures*, ACM/Springer Verlag Multimedia Systems Journal, Special Issue on QoS Architecture, 6 (3).

[56] Verma, S. (1998). *Multicast Routing and Resource Allocation for High-Speed Networks*, PhD Thesis, University of Toronto.

[57] Kamoun, F. and Kleinrock, L. (1979). *Stochastic Performance Evaluation of Hierarchical Routing for Large Networks*, Computer Networks, Vol.3, 337-353.

[58] Huitema, C. (1995). *Routing in the Internet*, Prentice Hall, New-Jersey, page 178.

[59] Maufer, T.A. (1998). *Deploying IP Multicast in the Enterprise*, Prentice Hall, New-Jersey, page 88.

10.0 Acronyms and Initialisms

AAP	Address Allocation Protocol
ACM	Association for Computing Machinery
AS	Autonomous System
ATM	Asynchronous Transfer Mode
BGMP	Border Gateway Multicast Protocol
BGP	Border Gateway Protocol
CBT	Core Based Trees
CDPD	Cellular Digital Packet Data
CGBT	Core Group Based Trees
CIDR	Classless Inter-Domain Routing
CPU	Central Processing Unit
DIS	Distributed Interactive Simulation
DR	Designated Router
DRP	Designated RP
DTCP	Dynamic Tunnel Configuration Protocol
DVMRP	Distance Vector Multicast Routing Protocol
DWR	Domain Wide Multicast Group Membership Report
EIGRP	Enhanced IGRP
GKDC	Group Key Distribution Centre
GKMP	Group Key Management Protocol
HPIM	Hierarchical PIM
IAB	Internet Architecture Board
IDMR	Inter-Domain Multicast Routing
IDPR	Inter-Domain Policy Routing
IDRP	Inter-Domain Routing Protocol
IESG	Internet Engineering Steering Group
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IGRP	Interior Gateway Routing Protocol
IP	Internet Protocol
IPMI	IP Multicast Initiative
IPSEC	Internet Security Protocol

IPX	Internet Packet Exchange
IS-IS	Intermediate System to Intermediate System
ISO	International Standards Organization
ISP	Internet Service Provider
LAN	Local Area Network
MAAS	Multicast Address Allocation Server
MALLOC	Multicast Address Allocation
MASC	Multicast Address-Set Claim
MBGP	Multicast BGP
MBONE	Multicast Backbone
MD	Multicast Domain
MDHCP	Multicast Address Allocation - Dynamic Host Configuration Protocol
MIGP	Multicast Interior Gateways Protocol
MIP	Multicast Internet Protocol
MOSPF	Multicast OSPF
NLSP	NetWare Link Services Protocol
NSFNET	National Science Foundation Network
OCBT	Ordered CBT
OSI	Open Systems Interconnection
OSPF	Open Shortest Path First
PIM	Protocol Independent Multicast
PIM-DM	PIM Dense Mode
PIM-SM	PIM Sparse Mode
PTMR	Policy Tree Multicast Routing
QoS	Quality of Service
QoSMIC	QoS Multicast Internet Protocol
QOSPF	QoS Path First Routing
RFC	Request for Comment
RIB	Routing Information Base
RIP	Routing Information Protocol
RP	Rendezvous Point
RPF	Reverse Path Forwarding
RSVP	Resource Reservation Protocol
SDRP	Source Demand Routing Protocol
SPF	Shortest Path First

TCP	Transmission Control Protocol
TOS	Type of Service
TRPB	Truncated Reverse Path Broadcasting
TTL	Time to Live
UDLR	Unidirectional Link Routing
VR	Virtual Router
WAN	Wide Area Network

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

COMMUNICATIONS RESEARCH CENTRE
3701 CARLING AVENUE, P.O. BOX 11490, STN H
OTTAWA, ONTARIO, CANADA, K2H8S2

2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable)

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

THE TRADE-OFFS OF MULTICAST ROUTING PROTOCOLS (U)

4. AUTHORS (Last name, first name, middle initial)

Bilodeau, Claude

5. DATE OF PUBLICATION (month and year of publication of document)

December 1999

6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)

xiv+50

6b. NO. OF REFS (total cited in document)

59

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

TECHNICAL REPORT

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

DEFENCE RESEARCH ESTABLISHMENT OTTAWA (DREO)
3701 CARLING AVENUE, OTTAWA, ONTARIO
K1A 0Z4

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)

WU#5cb14

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DREO TR 1999-119

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)

CRC-RP-99-004

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNLIMITED

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

During the past few years, several multicast routing protocols have emerged, which are competing to provide efficient mechanisms to deliver Internet Protocol (IP) traffic to groups of users scattered throughout the Internet. The multiplicity of experimental protocols and the absence of any well-established standardised protocol for multicast routing indicates that multicast routing has many solutions and that no one implementation can provide the most satisfactory characteristics in every situation.

This paper shows that much work is still needed to advance the state of the multicast routing technology. The main deficiencies of multicast routing protocols and their challenging design issues are illustrated by focusing on a few of the most popular multicast protocols being designed or experimented with today by the Internet Engineering Task Force (IETF).

Most of the multicast routing technology trade-offs analysed in the report apply to the global Internet in general while some are more specific to the tactical communication networks.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Multicast
Unicast
Routing Infrastructure
Routing Protocol
Policy Routing
QoS Routing
DVMRP
MOSPF
PIM
CBT
BGMP
Source-based trees
Share-trees

The Defence Research
and Development Branch
provides Science and
Technology leadership
in the advancement and
maintenance of Canada's
defence capabilities.

Leader en sciences et
technologie de la défense,
la Direction de la recherche
et du développement pour
la défense contribue
à maintenir et à
accroître les compétences
du Canada dans
ce domaine.

#512354

DEFENCE R&D DÉFENSE

www.crad.dnd.ca